FOR OFFICIAL USE ONLY

U.S. DEPARTMENT OF COMMERCE
Patent and Trademark Office

SEARCH REQUEST FORM

7-452

Examiner # **(Mandatory):** _____ Requester's Full Name: _____

Art Unit _____ Location (Bldg/Room#): _____ Phone (circle 305 306 308) _____

Serial Number: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

Title of Invention _____

Inventors (please provide full names): _____

_____

Earliest Priority Date: _____

Keywords (include any known synonyms registry numbers, explanation of initialisms):

Search Topic:

Please write detailed statement of the search topic, and the concept of the invention. Describe as specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples of relevant citations, authors, etc., if known. You may include a copy of the abstract and the broadcast or most relevant claim(s).

Point of Contact:
Alex Waclawiw
Technical Info. Specialist
CM1 12C14 Tel: 308-4491

STAFF USE ONLY

Searcher: _____
Searcher Phone #: _____
Searcher Location: _____
Date Picked Up: 7-24-99
Date Completed: 7-26-99
Clerical Prep Time: _____
Terminal Time: _____
Number of Databases: _____

Type of Search

1	N.A. Sequence
1	A.A. Sequence
_____	Structure (#)
_____	Bibliographic
_____	Litigation1
_____	Fulltext
_____	Procurement
_____	Other

Vendors (include cost where applicable)

_____	STN
_____	Questel/Orbit
_____	Lexis/Nexis
_____	WWW/Internet
_____	In-house sequence systems (list)
_____	Dialog
_____	Dr. Link
_____	Westlaw
X	Other (specify)

7-482

STIC-Biotech/ChemLib

From: Saoud, Christine
Sent: Saturday, July 17, 1999 7:12 AM
To: STIC-Biotech/ChemLib
Subject: sequence search

Importance: High

09/002,485

Please search SEQ ID NO:25 and 102 in the patent and commercial databases. Please keep any interference search separate.

Thank you,
christine saoud
A.U. 1646
10E03
305-7519

IMPORTANT INFORMATION ABOUT YOUR SEQUENCE SEARCH:

Compugen Sequence searching hardware and software explained:

This is the new sequence searching system that is currently being phased into as a replacement for the Maspar/Mpsrch platform. This system has been tested by both searchers and examiners, and has shown equivalent results to the Maspar system for the same databases. The results output format for all Compugen printed results are essentially the same **except for translations.**

Translation searching on Compugen explained:

The Compugen system utilizes Framesearch software for translations of proteins to nucleotides, and nucleotides to proteins. Some examiners have found these to be superior to the backtranslate software on Maspars.

FrameSearch searches a group of protein sequences for similarity to one or more nucleotide query sequences, or searches a group of nucleotide sequences for similarity to one or more protein query sequences. For each sequence comparison, the program finds an optimal alignment between the protein sequence and the corresponding codons on each the nucleotide sequence. Optimal alignments may include reading frame shifts. Please see any of the professional searching staff if you need assistance with this format.

File extensions for Compugen results transferred to floppy disks.

Compugen system search results will be delivered in one of two possible formats:

1. Standard concatenated files with .flp extension.
2. Compressed .zip files which decompressed yield two files as described below:

US08123456.cmr - Contains all commercial databases, may include Issued
US08123456.pen - Contains pending file results only

VERY IMPORTANT NOTE ABOUT PENDING FILE SEARCHES.

If your search contains file names with the following bolded extensions:
US08123456.**rap** US08123456.**rnp**

Do not leave this search in the case, during prosecution, or after the case issues, since it contains pending data which is confidential.

QUESTIONS? Contact any of the following:

Dilip Pandya, Chief, Information Branch, 308-4268

Professional searching staff:
John Dantzman (308-4488); Jan Delaval (308-4498); Mary Hale (308-4258); Barb O'Bryen (308-4291); David Schreiber (308-4292); Paula Sheppard (308-4499); Mark Spencer (308-4266); Beverly Shears (308-4994); Alex Waclawiw (308-4491).

GenCore version 4.5

OM protein - protein search, using sw model

Run on: July 24, 1999, 07:12:20 ; Search time 22.36 Seconds
(without alignments)
412.860 Million cell updates/sec

Title: US-09-002-485-25
Perfect score: 710
Sequence: 1 MNLWLLACLVAGFLGAWAPA.....NPISSSKRNVSLLISANSGL 150

Scoring table: PAM150

Searched: 201082 seqs, 61543640 residues

Database : SPTREMBL_10:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	98	13.8	96	11	Q9Z1X3	Q9z1x3 mus musculu
2	86.5	12.2	120	11	Q9Z1X0	Q9z1x0 mus musculu
3	83	11.7	94	12	Q98157	Q98157 kaposi's sa
4	74.5	10.5	91	6	O97919	O97919 bos taurus
5	70	9.9	1043	10	O82345	O82345 arabidopsis
6	67	9.4	372	11	Q9Z2D4	Q9z2d4 cavia porce
7	66.5	9.4	126	2	Q52196	Q52196 enterococcu
8	66	9.3	312	5	O45503	O45503 caenorhabdi
9	64.5	9.1	818	2	Q9ZGE3	Q9zge3 heliobacill
10	64.5	9.1	392	11	O55145	O55145 rattus norv
11	64.5	9.1	97	11	Q9Z318	Q9z318 cavia porce
12	64	9.0	992	10	O65440	O65440 arabidopsis
13	64	9.0	395	11	O35933	O35933 mus musculu
14	64	9.0	395	11	O35188	O35188 mus musculu
15	63.5	8.9	935	13	O42596	O42596 xenopus lae
16	63	8.9	393	2	O50532	O50532 streptomyce
17	62.5	8.8	255	2	O83255	O83255 treponema p
18	62.5	8.8	80	4	Q14745	Q14745 homo sapien
19	62.5	8.8	207	6	O77513	O77513 oryctolagus
20	62.5	8.8	473	11	O55197	O55197 rattus norv
21	62	8.7	510	3	O93992	O93992 candida alb
22	61.5	8.7	596	2	Q9Z493	Q9z493 flavobacter
23	61.5	8.7	308	4	Q15071	Q15071 homo sapien
24	61	8.6	302	1	O28822	O28822 archaeoglob
25	61	8.6	129	3	O59860	O59860 lentinula e
26	61	8.6	147	8	Q02705	Q02705 podospora a
27	60.5	8.5	2948	5	Q93785	Q93785 caenorhabdi
28	60.5	8.5	164	10	Q43659	Q43659 triticum ae
29	60.5	8.5	534	12	Q66643	Q66643 equine herp
30	60.5	8.5	95	12	Q98158	Q98158 kaposi's sa
31	60	8.5	320	2	Q55350	Q55350 synechococc
32	60	8.5	151	2	O84177	O84177 chlamydia t
33	60	8.5	472	2	Q9ZG88	Q9zg88 caulobacter
34	60	8.5	1389	10	O24564	O24564 zea mays (m
35	60	8.5	2629	11	P97499	P97499 mus musculu
36	60	8.5	1608	11	Q9Z286	Q9z286 rattus norv
37	60	8.5	1811	12	O36184	O36184 plautia sta
38	59.5	8.4	846	2	O05442	O05442 mycobacteri
39	59.5	8.4	177	2	O83563	O83563 treponema p
40	59.5	8.4	78	2	O86357	O86357 mycobacteri
41	59.5	8.4	265	2	O50965	O50965 borrelia bu
42	59.5	8.4	931	5	O18116	O18116 caenorhabdi
43	59	8.3	308	2	P96819	P96819 mycobacteri
44	59	8.3	316	2	Q60064	Q60064 vibrio chol
45	59	8.3	316	2	Q60118	Q60118 vibrio chol

ALIGNMENTS

RESULT 1
Q9Z1X3
ID Q9Z1X3 PRELIMINARY; PRT; 96 AA.
AC Q9Z1X3;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE MACROPHAGE INFLAMMATORY PROTEIN 3 ALPHA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA MCCOLL S.R., CARNEY T., NEOTE K.S.;
RT "Cloning and Characterization of a murine homolog of macrophage
RT inflammatory protein 3 alpha.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF099052; AAC78680.1; -.
SQ SEQUENCE 96 AA; 10755 MW; F5A0C651 CRC32;

Query Match 13.8%; Score 98; DB 11; Length 96;
Best Local Similarity 33.3%; Pred. No. 0.00016;
Matches 30; Conservative 21; Mismatches 27; Indels 12; Gaps 6;

```
Qy      13 FLG-AWAPAVH----AQGVFEDCCLAY-HYPI-GWAVLRRAWTYRIQEVSGSCNLPAAIF 65
           ||: ||    |    |::   ||||:| : |: ::|::     :  |    :|:: | ||
Db      10 FLALAWVLLAHLCSQAEASNYDCCLSYIQTPLPSRAIV----GFTRQMADEACDINAIIF 65

Qy      66 YLPKRHRKVCGNPKSREVQRAMKLLDARNK 95
                |::: ||::||   | ||::||  | |
Db      66 H-TKKRKSVCADPKQNWVKRAVNLLSLRVK 94
```

RESULT 2
Q9Z1X0
ID Q9Z1X0 PRELIMINARY; PRT; 120 AA.
AC Q9Z1X0;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE CC CHEMOKINE ALP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RA HOU Y.H., CHRISTOPHERSON K., HROMAS R.;
RT "ALP, a novel highly divergent CC chemokine.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF099931; AAD04163.1; -.

```
SQ   SEQUENCE   120 AA;  13441 MW;  A040A17E CRC32;

  Query Match          12.2%;  Score 86.5;  DB 11;  Length 120;
  Best Local Similarity   35.1%;  Pred. No. 0.0048;
  Matches   20;  Conservative   12;  Mismatches   22;  Indels    3;  Gaps    2;

Qy      30 CCLA-YHYPIGWAVLRRAWTYRIQEVSGSCNLPAAIFYLPKRHRKVCGNPKSREVQR 85
           ||   |: |:   :|||     :|| | |:| | :: |::  | || :| :| : |
Db      34 CCTQLYRQPLPSRLLRRIVHMELQEADGDCHLQAVVLHLAR--RSVCVHPQNRSLAR 88


RESULT   3
Q98157
ID   Q98157      PRELIMINARY;      PRT;    94 AA.
AC   Q98157;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   VMIP-1B.
OS   Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8),
OS   and Kaposi's sarcoma-associated herpesvirus (KSHV)
OS   (Human herpesvirus 8).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Rhadinovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   NICHOLAS J., RUVOLO V.R., BURNS W.H., SANDFORD G., WAN X., CIUFO D.,
RA   HENDRICKSON S., GUO H.G., HAYWARD G.S., REITZ M.S.;
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97094384.
RA   MOORE P.S., BASHOFF C., WEISS R.A., CHANG Y.;
RT   "Molecular mimicry of human cytokine and cytokine response pathway
RT   genes by KSHV.";
RL   Science 274:1739-1744(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97121480.
RA   RUSSO J.J., BOHENZKY R.A., CHIEN M.C., CHEN J., YAN M., MADDALENA D.,
RA   PARRY J.P., PERUZZI D., EDELMAN I.S., CHANG Y., MOORE P.S.;
RT   "Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus
RT   (HHV8).";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:14862-14867(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   RUSSO J.J., BOHENZKY R.A., CHIEN M.C., CHEN J., YAN M., MADDALENA D.,
RA   PARRY J.P., PERUZZI D., EDELMAN I.S., CHANG Y., MOORE P.S.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   RUSSO J.J., BOHENZKY R.A., CHIEN M.C., CHEN J., YAN M., MADDALENA D.,
RA   PARRY J.P., PERUZZI D., EDELMAN I.S., CHANG Y., MOORE P.S.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97296220.
RA   NEIPEL F., ALBRECHT J.C., FLECKENSTEIN B.;
RT   "Cell-homologous genes in the Kaposi's sarcoma-associated
RT   rhadinovirus human herpesvirus 8: determinants of its
RT   pathogenicity?";
RL   J. Virol. 71:4187-4192(1997).
RN   [7]
RP   SEQUENCE FROM N.A.
RA   SUN R., LIN S.-F., MILLER G.;
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U67775; AAB61702.1; -.
DR   EMBL; U75698; AAC57093.1; -.
DR   EMBL; U93872; AAB62642.1; -.
DR   EMBL; U71365; AAC34941.1; -.
DR   PFAM; PF00048; il8; 1.
KW   Hypothetical protein.

SQ   SEQUENCE   94 AA;  10486 MW;  91312200 CRC32;

  Query Match          11.7%;  Score 83;  DB 12;  Length 94;
  Best Local Similarity   31.8%;  Pred. No. 0.0098;
  Matches   28;  Conservative   18;  Mismatches   26;  Indels   16;  Gaps    7;

Qy       6 LACLVAG--FLGAWAPAVHAQGVFEDCCLAYHYPIGWAVLRRAWTYRIQEVSGSCNLPAA 63
           | || :|  : ::|  :       : |||:|:      ||  :| |   ::   |: |:
Db      15 LLCLQSGDTLGASWHRP------DKCCLGYQKRPLPQVLLSSW-YPTSQL---CSKPGV 63

Qy      64 IFYLPKRHRKVCGNPKSRE-VQRAMKLL 90
           || | || | ||:: ||:: | : |  |
Db      64 IF-LTKRGRQVCAD-KSKDWVKKLMQQL 89


RESULT   4
O97919
ID   O97919      PRELIMINARY;      PRT;    91 AA.
AC   O97919;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   RANTES PROTEIN.
GN   RANTES.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae;
OC   Bovinae; Bos.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE;
RA   AUST G., THAMM B., ROST A.K.;
RT   "Cloning of Bos taurus RANTES mRNA.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ007043; CAA07430.1; -.
DR   PROSITE; PS00472; SMALL_CYTOKINES_CC; 1.
SQ   SEQUENCE   91 AA;  10073 MW;  1CD014C6 CRC32;

  Query Match          10.5%;  Score 74.5;  DB 6;  Length 91;
  Best Local Similarity   29.2%;  Pred. No. 0.1;
  Matches   19;  Conservative   21;  Mismatches   14;  Indels   11;  Gaps    4;

Qy      30 CCLAYHYPIGWAVLRRAWTYRIQE---VSGSCNLPAAIFYLPKRHRKVCGNPKSREVQRA 86
           ||:||   |:::: |     ::||    |: |:::| :| : :: | ||:||  : |:
Db      33 CCFAY---ISRPLPRT----HVQEYFYTSSKCSMAAVVF-ITRKKRQVCANPEKKWVREY 84

Qy      87 MKLLD 91
           :: |:
Db      85 INALE 89


RESULT   5
O82345
ID   O82345      PRELIMINARY;      PRT;  1043 AA.
AC   O82345;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   T3F17.11 PROTEIN.
GN   T3F17.11.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   ROUNSLEY S.D., LIN X., KAUL S., SHEA T.P., FUJII C.Y., MASON T.M.,
RA   SHEN M., RONNING C.M., FRASER C.M., SOMERVILLE C.R., VENTER J.C.;
RT   "Arabidopsis thaliana chromosome II BAC T3F17 genomic sequence.";
```

```
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC005397; AAC62882.1; -.
SQ   SEQUENCE   1043 AA;  116758 MW;  9632EF1A CRC32;


  Query Match             9.9%;  Score 70;  DB 10;  Length 1043;
  Best Local Similarity   19.0%;  Pred. No. 4.6;
  Matches   22;  Conservative   25;  Mismatches   29;  Indels   40;  Gaps    5;

Qy       7 ACLVAGFLGAWAPAVHAQGVFEDCCL---AYH--------------YPIGWAVLRRAWT 48
           :|: : | : |:|   :|  : :||:   |:|              :|:|     ::|
Db      42 SCVHGNFPAYWPPCYPPQVPYHQCCMNRSAFHPPHASYAPSCYVHPPFPVGY----QPWF 97

Qy      49 YRIQEVSGS----------CNLPAAIFYLPKRHRKVCGNPKSREVQRAMKLLDARN 94
                :|:|           |:|        |: | |  :    |:::: :|  |:
Db      98 DVEKDVPGKHHCGKCSSQMCDL--------KKDRGVVIEEHEPEIEKGEAVLPVRS 145


RESULT   6
Q9Z2D4
ID   Q9Z2D4      PRELIMINARY;      PRT;   372 AA.
AC   Q9Z2D4;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PANCREATIC POLYPEPTIDE RECEPTOR Y4.
OS   Cavia porcellus (Guinea pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 99017377.
RA   ERIKSSON H., BERGLUND M.M., HOLMBERG S.K., KAHL U., GEHLERT D.R.,
RA   LARHAMMAR D.;
RT   "The cloned guinea pig pancreatic polypeptide receptor Y4 resembles
RT   more the human Y4 than does the rat Y4.";
RL   Regul. Pept. 75:29-37(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   SHARMA P.S., HOLMBERG S.K., ERIKSSON H., BECK-SICKINGER A.G.,
RA   GRUNDEMAR L., LARHAMMAR D.;
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF072822; AAD13144.1; -.
KW   Receptor.
SQ   SEQUENCE   372 AA;  41774 MW;  4CED6956 CRC32;


  Query Match             9.4%;  Score 67;  DB 11;  Length 372;
  Best Local Similarity   20.0%;  Pred. No. 3.5;
  Matches   28;  Conservative   23;  Mismatches   37;  Indels   52;  Gaps    5;

Qy       3 LWLLACLVA--------------------GFL-------GAWAPAVHAQGVFEDCCLAY 34
           :||:||:::                   :||       :| |  | : |:    | :
Db     163 IWLIACVLSLPFLANSILENVFHKNHSKALAFLEDKVACKESW-PLGHHRVVYTTFLLLF 221

Qy      35 HYPIGWAV-----LRRAWTYRIQEVSGSCNLPAAIFYLPKRHRKVCGNPKSREVQRAMKL 89
           :| |  |       || :|  | |                  | :| :::: :: |   :
Db     222 QYCIPLAFILVCYLRISWRLRRQG-----------------RVLCKSSRAGQMKRINGV 263

Qy      90 LDARNKVFAKLRHNTQTFQA 109
           | |    || |:   : | :
Db     264 LVAMVGAFAVLWLPLHVFNS 283


RESULT   7
Q52196
ID   Q52196      PRELIMINARY;      PRT;   126 AA.
AC   Q52196;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   TRAE.
GN   TRAE.
OS   Enterococcus faecalis (Streptococcus faecalis).
OG   Plasmid pPD1.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC   Enterococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96032393.
RA   NAKAYAMA J., YOSHIDA K., KOBAYASHI H., ISOGAI A., CLEWELL D.,
RA   SUZUKI A.;
RT   "Cloning and characterization of a region of Enterococcus faecalis
RT   plasmid pPD1 encoding pheromone inhibitor (ipd), pheromone
RT   sensitivity (traC), and pheromone shutdown (traB) genes.";
RL   J. Bacteriol. 177:5567-5573(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96032394.
RA   FUJIMOTO S., TOMITA H., WAKAMATSU E., TANIMOTO K., IKE Y.;
RT   "Physical mapping of the conjugative bacteriocin plasmid pPD1 of
RT   Enterococcus faecalis and identification of the determinant related
RT   to the pheromone response.";
RL   J. Bacteriol. 177:5574-5581(1995).
DR   EMBL; D78016; BAA11199.1; -.
KW   Plasmid.
SQ   SEQUENCE   126 AA;  14896 MW;  D5A80A91 CRC32;


  Query Match             9.4%;  Score 66.5;  DB 2;  Length 126;
  Best Local Similarity   19.3%;  Pred. No. 1.3;
  Matches   28;  Conservative   28;  Mismatches   38;  Indels   51;  Gaps    7;

Qy      34 YHYPIGWAV---LRRAW----------------TYRIQEVSGSC-NLPAAIFYLP----- 69
           | :|: | :   :|::|                |||:  :|:|  |    |||:
Db       2 YMFPVFWFLSLNFRKGWCSRMKQKVTGTMILQKVYRIHTKTGNCYVLENERFYLAVQVLE 61

Qy      69 ------KRHRKVCGNPKSREVQRAMKLLDARNKVFAKLRHNTQTFQAGPHAVKKLSSGNS 122
                |: | :       | :|:          ||:    ::   : :    |::
Db      62 HGKMIWKKYREI----PQMGVTKAL-----------LRQIISCYDKDGNKIIERKKGST 105

Qy     123 KLSSSKFSNPISSSKRNVSLLISAN 147
           ||:      |:   ::: | |:||:
Db     106 KLTLV----PVKKVNKKRSQLVSAS 126


RESULT   8
O45503
ID   O45503      PRELIMINARY;      PRT;   312 AA.
AC   O45503;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   F39B2.11 PROTEIN.
GN   F39B2.11.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   DOBSON R.;
RL   Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
```

```
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z92834; CAB07391.1; -.
SQ   SEQUENCE   312 AA;  35319 MW;  B98796E4 CRC32;


  Query Match             9.3%;  Score 66;  DB 5;  Length 312;
  Best Local Similarity   24.3%;  Pred. No. 3.9;
  Matches   27;  Conservative   25;  Mismatches   19;  Indels   40;  Gaps    7;

Qy     30 CCLAYH-YPIGWAVLRRAWT----------YRIQEVSGSCNLPAAIFYLPKRHRKVCGNP 78
          | |  : ||   ||::  ||          |:    ::  ::|  ::|| ||:::|
Db     98 CYLHHNLYP---AVMHTFWTDELNYNTVTQYWY---ASHLHFPYNLYYLEKRRKK----- 147

Qy     79 KSREVQRAMKLLDARNKVFAKLRHNTQTFQAGPHAVKKLSSGNSKLSSSKF 129
             |::|| ::|        :|: :  :  |:: ||   :||: :||
Db    147 -------ALRLLAGKN--------DTEILKEAFMALNTLS---TKLGDNKF 179


RESULT   9
Q9ZGE3
ID   Q9ZGE3      PRELIMINARY;      PRT;   818 AA.
AC   Q9ZGE3;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   STAGE II SPORULATION PROTEIN E SP2E.
GN   SP2E.
OS   Heliobacillus mobilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Heliobacterium group; Heliobacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   XIONG J., INOUE K., BAUER C.E.;
RT   "Tracking molecular evolution of photosynthesis by characterization
RT   of a major photosynthesis gene cluster from Heliobacillus mobilis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:14851-14856(1998).
DR   EMBL; AF080002; AAC84035.1; -.
SQ   SEQUENCE   818 AA;  88722 MW;  A5EE782A CRC32;


  Query Match             9.1%;  Score 64.5;  DB 2;  Length 818;
  Best Local Similarity   22.3%;  Pred. No. 16;
  Matches   31;  Conservative   30;  Mismatches   31;  Indels   47;  Gaps    8;

Qy      1 MNLWLLACLVAGFLGAWA-PAVHA-------------------QGVFEDCC---LAYHY 36
          | |  |: |:|| : :|| |:| :                   |::||      : :
Db    267 MALFALSGLLAGVFRPWAKPGVLGGFFLGHLLLSIYMGGSREVHQALFEATVAGTVLFFW 326

Qy     37 PIGW-AVLRRAWTYRIQEVSGSCNLPAAIFYLPKRHRKVCGNPKSREVQRAMKLLDARNK 95
          |::|  ||||    ||    ::   : | : | : ::|: |:             |:   :
Db    327 PVSWLQVLRR----RITKRDAA---ANAALELENNRKKLAGK------------LEEMGR 367

Qy     96 VFAKLRHNTQTFQAGPHAV 114
          :|: |   : ||:  :::|
Db    368 MFGQL---AVTFDEVSQPV 383


RESULT  10
O55145
ID   O55145      PRELIMINARY;      PRT;   392 AA.
AC   O55145;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   CHEMOKINE CX3C PRECURSOR (FRACTALKINE/NEUROACTIN).
GN   ACC1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98393742.
RA   HARRISON J.K., JIANG Y., CHEN S., XIA Y., MACIEJEWSKI D.,
RA   MCNAMARA R.K., STREIT W.J., SALAFRANCA M.N., ADHIKARI S.,
RA   THOMPSON D.A., BOTTI P., BACON K.B., FENG L.;
RT   "Role for neuronally derived fractalkine in mediating interactions
RT   between neurons and CX3CR1-expressing microglia.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:10896-10901(1998).
RN   [2]
RP   SEQUENCE OF 203-334 FROM N.A.
RA   SCHWAEBLE W., STOVER C.M., TRINDER P.K.E., LINNINGTON C., IGLESIAS A.,
RA   LYNCH N.J., WEIHE E., SCHAFER M.K.H.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF030358; AAC33834.1; -.
DR   EMBL; Y16813; CAA76404.1; -.
KW   Signal.
FT   SIGNAL        1     24       POTENTIAL.
SQ   SEQUENCE   392 AA;  42162 MW;  8A36DF93 CRC32;


  Query Match             9.1%;  Score 64.5;  DB 11;  Length 392;
  Best Local Similarity   30.8%;  Pred. No. 7.5;
  Matches   20;  Conservative   16;  Mismatches   20;  Indels    9;  Gaps    2;

Qy     63 AIFYLPKRHRKVCGNPKSREVQRAMKLLDARNKVFAKLRHNTQTFQAGPHAVKKLSSGNS 122
          ||    ::||  |::|| : || ||| || :    | | :|   |:      |:: :  :
Db     62 AIILETRQHRHFCADPKEKWVQDAMKHLDHQT---AALTRNGGKFE------KRVDNVTP 112

Qy    123 KLSSS 127
          :::|:
Db    113 RITSA 117


RESULT  11
Q9Z318
ID   Q9Z318      PRELIMINARY;      PRT;    97 AA.
AC   Q9Z318;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   MONOCYTE CHEMOATTRACTANT PROTEIN-3 (MCP-3).
OS   Cavia porcellus (Guinea pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LUNG;
RA   ASANO K., NAKAMURA M., OGUMA T., FUKUNAGA K., MATSUBARA H.,
RA   ISHIZAKA A., YAMAGUCHI K., KANAZAWA M.;
RT   "Differential expression of CC chemokines in guinea pig lungs during
RT   an allergic inflammation.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB014340; BAA36456.1; -.
DR   PROSITE; PS00472; SMALL_CYTOKINES_CC; 1.
SQ   SEQUENCE   97 AA;  11159 MW;  8C35C708 CRC32;


  Query Match             9.1%;  Score 64.5;  DB 11;  Length 97;
  Best Local Similarity   26.7%;  Pred. No. 1.7;
  Matches   28;  Conservative   22;  Mismatches   24;  Indels   31;  Gaps    6;

Qy      5 LLACLV------AGFLGAWAPAVHAQGV-FEDCCLAYHYPIGWAVLRRAWTYRIQEV--- 55
            :| ||:      :::| |     :::||    ||           :|:   |:| :
Db      6 VLLCLLLTAANFSSLLLA-----QPDGVNISTCC----------YKRSQRIRVQRLESY 49

Qy     55 ----SGSCNLPAAIFYLPKRHRKVCGNPKSREVQRAMKLLDARNK 95
               |: |   | ||   | :| :|::|| : || :|| :| ::|
Db     50 TRITSSKCPWQAVIF-KTKFNREICADPKQQWVQDSMKYIDKKSK 93


RESULT  12
O65440
ID   O65440      PRELIMINARY;      PRT;   992 AA.
AC   O65440;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
```

```
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   CLV1 RECEPTOR KINASE LIKE PROTEIN.
GN   F1C12.190.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BEVAN M., TERRYN N., ARDILES W., BUYSSHAERT C., DASSEVILLE R.,
RA   DE CLERCK R., DE KEYSER A., NEYT P., ROUZE P., VAN DEN DAELE H.,
RA   VILLAROEL R., GIELEN J., VAN MONTAGU M., BANCROFT I., MEWES H.W.,
RA   MAYER K., SCHUELLER C.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU ARABIDOPSIS SEQUENCING PROJECT;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL022224; CAA18252.1; -.
DR   PFAM; PF00560; LRR; 11.
DR   PFAM; PF00069; pkinase; 1.
SQ   SEQUENCE   992 AA;  110283 MW;  C795E06D CRC32;


  Query Match             9.0%;  Score 64;  DB 10;  Length 992;
  Best Local Similarity   24.8%;  Pred. No. 23;
  Matches   26;  Conservative   25;  Mismatches   28;  Indels   26;  Gaps    4;

Qy      47 WTYRIQEVSGSCNLPAAIFYLPKRHRKVCGNPKSREVQRAMKLLDARNKVFAKLRHNTQT 106
           | :|: :   : :|| :: |||:                : ||: :|:    |  :
Db     418 WRFRLGQNFLTSKLPKGLIYLPN----------------LSLLELQNNF---LTGEIPE 457

Qy     107 FQAGPHAVKKLSSGNSKLSSSKFSNPISSSKRNVS----LLISAN 147
           :||      |:    :||:::::| ||::| ||:     ||::||
Db     458 EEAGNAQFSSLT--QINLSNNRLSGPIPGSIRNLRSLQILLLGAN 500


RESULT  13
O35933
ID   O35933      PRELIMINARY;      PRT;   395 AA.
AC   O35933;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   FRACTALKINE.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=BRAIN;
RA   ROSSI D., HARDIMAN G., COPELAND N., GILBERT D.J., JENKINS N.,
RA   ZLOTNIK A., BAZAN J.F.;
RL   Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U92565; AAB71763.1; -.
DR   PFAM; PF00048; il8; 1.
SQ   SEQUENCE   395 AA;  42040 MW;  3997A113 CRC32;


  Query Match             9.0%;  Score 64;  DB 11;  Length 395;
  Best Local Similarity   29.7%;  Pred. No. 8.6;
  Matches   22;  Conservative   21;  Mismatches   27;  Indels    4;  Gaps    2;

Qy      70 RHRKVCGNPKSREVQRAMKLLDARNKVFAKLRHNTQTFQAGPHAVK-KLSSGNSKLSSSK 128
           :||: |::|| : || ||| || :   ::|   |   |:     |   :: :   ||:|
Db      69 QHRRFCADPKEKWVQDAMKHLDHQAAALTK---NGGKFEKRVDNVTPGITLATRGLSPSA 125

Qy     129 FSNPISSSKRNVSL 142
           :::| |::  :::|
Db     126 LTKPESATLEDLAL 139


RESULT  14
O35188
ID   O35188      PRELIMINARY;      PRT;   395 AA.
AC   O35188;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   NEUROTACTIN.
GN   SCYD1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97320499.
RA   PAN Y., CLARE L., HONG Z., DOLICH S., DEEDS J., GONZALO J., VATH J.,
RA   GOSSELIN M., MA J., DUSSAULT B., WOOLF B., ALPERIN A., CULPEPPER J.,
RA   GUTIERREZ-RAMOS J.C., GEARING D.;
RT   "Neurotactin, a membrane-anchored chemokine upregulated in brain
RT   inflammation.";
RL   Nature 387:611-617(1997).
DR   EMBL; AF010586; AAB66331.1; -.
DR   MGD; MGI:1097153; SCYD1.
DR   PFAM; PF00048; il8; 1.
SQ   SEQUENCE   395 AA;  42098 MW;  E3CD0612 CRC32;


  Query Match             9.0%;  Score 64;  DB 11;  Length 395;
  Best Local Similarity   29.7%;  Pred. No. 8.6;
  Matches   22;  Conservative   21;  Mismatches   27;  Indels    4;  Gaps    2;

Qy      70 RHRKVCGNPKSREVQRAMKLLDARNKVFAKLRHNTQTFQAGPHAVK-KLSSGNSKLSSSK 128
           :||: |::|| : || ||| || :   ::|   |   |:     |   :: :   ||:|
Db      69 QHRRFCADPKEKWVQDAMKHLDHQAAALTK---NGGKFEKRVDNVTPGITLATRGLSPSA 125

Qy     129 FSNPISSSKRNVSL 142
           :::| |::  :::|
Db     126 LTKPESATLEDLAL 139


RESULT  15
O42596
ID   O42596      PRELIMINARY;      PRT;   935 AA.
AC   O42596;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   METALLOPROTEASE-DISINTEGRIN MDC11B (METALLOPROTEASE/DISINTEGRIN
DE   XMDC11.2).
GN   MDC11B.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC   Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae;
OC   Xenopus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   CAI H., KRAETZSCHMAR J., ALFANDARI D., HUNNICUTT G., BLOBEL C.P.;
RT   "Neural crest-specific and general expression of distinct
RT   metalloprotease-disintegrins in early Xenopus laevis development.";
RL   Dev. Biol. 0:0-0(1998).
RN   [2]
RP   SEQUENCE OF 464-511 FROM N.A.
RC   TISSUE=TESTIS;
RX   MEDLINE; 97349132.
RA   SHILLING F.M., KRATZSCHMAR J., CAI H., WESKAMP G., GAYKO U.,
RA   LEIBOW J., MYLES D.G., NUCCITELLI R., BLOBEL C.P.;
RT   "Identification of metalloprotease/disintegrins in Xenopus laevis
RT   testis with a potential role in fertilization.";
RL   Dev. Biol. 186:155-164(1997).
DR   EMBL; AF032383; AAC61847.1; -.
DR   EMBL; U78188; AAB87148.1; -.
DR   PROSITE; PS00427; DISINTEGRINS; 1.
```

```
KW   Hydrolase; Metalloprotease; Protease.
SQ   SEQUENCE   935 AA;  104161 MW;  FF435012 CRC32;


  Query Match             8.9%;  Score 63.5;  DB 13;  Length 935;
  Best Local Similarity   20.2%;  Pred. No. 25;
  Matches   26;  Conservative   36;  Mismatches   34;  Indels   33;  Gaps    5;

Qy      19 PAVHAQGVFEDCCLAYHYPIGWAV---------LRRAWTYRIQEVSGSCNLPAAIFYLPK 69
           :|: ::     | :: ||:|:  :         | ::|: || :               |
Db     804 SAILSHFQISACSIP-HYSISQNISLFCRRSNGLSHSWSERIPD--------------TK 848

Qy      70 RHRKVCGNPKSREVQRAMKLLDARNKVFAKLRHNTQTFQAGPHAVKKLSSGNSKLSSSKF 129
           :   || | ::|   :  ::  :|:|: :|       |:: :::   ||:::   |:|
Db     849 HVSDVCENGRPRSNSWQGNVTSSRKKLRGK------RFRPRSNSTETLSPAK---SPSSS 899

Qy     130 SNPISSSKR 138
           : :|:|:|
Db     900 TGSIASSRR 908


Search completed: July 24, 1999, 10:27:22
Job time: 11702 sec
```

OM protein - protein search, using sw model

Run on: July 24, 1999, 07:09:12 ; Search time 13.41 Seconds
(without alignments)
316.200 Million cell updates/sec

Title: US-09-002-485-25
Perfect score: 710
Sequence: 1 MNLWLLACLVAGFLGAWAPA.....NPISSSKRNVSLLISANSGL 150

Scoring table: PAM150

Searched: 77977 seqs, 28268293 residues

Database : SwissProt_37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	698	98.3	150	1	TECK_HUMAN	O15444 homo sapien
2	312	43.9	144	1	TECK_MOUSE	O35903 mus musculu
3	98	13.8	122	1	MI1G_MOUSE	P51670 mus musculu
4	89	12.5	121	1	MIP3_HUMAN	P55773 homo sapien
5	81.5	11.5	94	1	TARC_HUMAN	Q92583 homo sapien
6	81	11.4	116	1	C10_MOUSE	P27784 mus musculu
7	80	11.3	92	1	MI1A_RAT	P50229 rattus norv
8	79	11.1	89	1	MIP4_HUMAN	P55774 homo sapien
9	78.5	11.1	97	1	EOTA_HUMAN	P51671 homo sapien
10	76.5	10.8	113	1	MIP5_HUMAN	Q16663 homo sapien
11	76.5	10.8	91	1	SISD_HUMAN	P13501 homo sapien
12	76	10.7	96	1	MI3A_HUMAN	P78556 homo sapien
13	73.5	10.4	98	1	MI3B_HUMAN	Q99731 homo sapien
14	73	10.3	92	1	MI1A_MOUSE	P10855 mus musculu
15	72.5	10.2	114	1	LTN_MOUSE	P47993 mus musculu
16	72	10.1	93	1	CCC1_HUMAN	Q16627 homo sapien
17	72	10.1	109	1	CCC3_HUMAN	Q13954 homo sapien
18	72	10.1	114	1	LTN_RAT	P51672 rattus norv
19	71	10.0	92	1	MI1B_MOUSE	P14097 mus musculu
20	70	9.9	96	1	I309_HUMAN	P22362 homo sapien
21	69.5	9.8	120	1	MCP1_CAVPO	Q08782 cavia porce
22	69.5	9.8	99	1	MCP2_HUMAN	P80075 homo sapien
23	69	9.7	99	1	MCP2_BOVIN	Q09141 bos taurus
24	66.5	9.4	389	1	RF1M_CAEEL	O44568 caenorhabdi
25	66.5	9.4	91	1	SISD_MOUSE	P30882 mus musculu
26	66.5	9.4	92	1	SISD_RAT	P50231 rattus norv
27	66	9.3	99	1	MCP2_PIG	P49873 sus scrofa
28	65.5	9.2	92	1	MI1B_RABIT	P46632 oryctolagus
29	65.5	9.2	91	1	SISD_CAVPO	P97272 cavia porce
30	65	9.2	96	1	EOTA_CAVPO	P80325 cavia porce
31	65	9.2	92	1	MI1B_HUMAN	P13236 h macrophag
32	64.5	9.1	92	1	MI1A_HUMAN	P10147 homo sapien
33	64.5	9.1	608	1	UL27_HCMVA	P16763 human cytom
34	63.5	8.9	114	1	LTN_HUMAN	P47992 homo sapien
35	63	8.9	148	1	MCP1_MOUSE	P10148 mus musculu
36	63	8.9	3655	1	YAMB_SCHPO	Q10064 schizosacch
37	62.5	8.8	93	1	MI10_HUMAN	P16619 homo sapien
38	62.5	8.8	92	1	MI1B_RAT	P50230 rattus norv
39	62.5	8.8	246	1	RL7_DICDI	P11874 dictyosteli
40	62.5	8.8	92	1	SISF_MOUSE	P10146 mus musculu
41	62	8.7	622	1	DCTB_RHILE	P10047 rhizobium l
42	61.5	8.7	370	1	5H5B_MOUSE	P31387 mus musculu
43	61.5	8.7	443	1	CGLH_XANMA	P37126 xanthomonas
44	61.5	8.7	389	1	OXYR_HUMAN	P30559 homo sapien
45	61.5	8.7	389	1	OXYR_MACMU	P56494 macaca mula

ALIGNMENTS

```
RESULT   1
TECK_HUMAN
ID   TECK_HUMAN     STANDARD;      PRT;   150 AA.
AC   O15444;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   CHEMOKINE TECK PRECURSOR (THYMUS EXPRESSED CHEMOKINE).
GN   SCYA25 OR TECK.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=THYMUS;
RX   MEDLINE; 97429775.
RA   VICARI A.P., FIGUEROA D.J., HEDRICK J.A., FOSTER J.S., SINGH K.P.,
RA   MENON S., COPELAND N.G., GILBERT D.J., JENKINS N.A., BACON K.B.,
RA   ZLOTNIK A.;
RT   "TECK: a novel CC chemokine specifically expressed by thymic
RT   dendritic cells and potentially involved in T cell development.";
RL   IMMUNITY 7:291-301(1997).
CC   -!- FUNCTION: POTENTIALLY INVOLVED IN T CELL DEVELOPMENT. RECOMBINANT
CC       TECK SHOWS CHEMOTACTIC ACTIVITY ON THYMOCYTES, MACROPHAGES, THP-1
CC       CELLS, AND DENDRITICS CELLS BUT IS INACTIVE ON PERIPHERAL BLOOD
CC       LYMPHOCYTES AND NEUTROPHILS.
CC   -!- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED BY THYMIC DENDRITIC
CC       CELLS. HIGH LEVELS IN THYMUS AND SMALL INTESTINE.
CC   -!- SIMILARITY: BELONGS TO THE INTERCRINE BETA FAMILY (SMALL CYTOKINE
CC       C-C) (CHEMOKINE CC).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U86358; G2388627; -.
DR   MIM; 602565; -.
DR   PROSITE; PS00472; SMALL_CYTOKINES_CC; FALSE_NEG.
KW   CYTOKINE; CHEMOTAXIS; INFLAMMATORY RESPONSE; SIGNAL.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24    150       CHEMOKINE TECK.
FT   DISULFID     30     58       BY SIMILARITY.
FT   DISULFID     31     75       BY SIMILARITY.
SQ   SEQUENCE   150 AA;  16639 MW;  C712E950 CRC32;


  Query Match             98.3%;  Score 698;  DB 1;  Length 150;
  Best Local Similarity   98.0%;  Pred. No. 2.7e-76;
  Matches  147;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 MNLWLLACLVAGFLGAWAPAVHAQGVFEDCCLAYHYPIGWAVLRRAWTYRIQEVSGSCNL 60
           |||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db       1 MNLWLLACLVAGFLGAWAPAVHTQGVFEDCCLAYHYPIGWAVLRRAWTYRIQEVSGSCNL 60

Qy      61 PAAIFYLPKRHRKVCGNPKSREVQRAMKLLDARNKVFAKLRHNTQTFQAGPHAVKKLSSG 120
           |||||||||||||||||||||||||||||||||||||||||:|| ||||||||||||||||
Db      61 PAAIFYLPKRHRKVCGNPKSREVQRAMKLLDARNKVFAKLHHNMQTFQAGPHAVKKLSSG 120

Qy     121 NSKLSSSKFSNPISSSKRNVSLLISANSGL 150
           |||||||||||||||||||||||||||||||
Db     121 NSKLSSSKFSNPISSSKRNVSLLISANSGL 150
```

```
RESULT   2
TECK_MOUSE
ID   TECK_MOUSE     STANDARD;      PRT;   144 AA.
AC   O35903;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   CHEMOKINE TECK PRECURSOR (THYMUS EXPRESSED CHEMOKINE).
GN   SCYA25 OR TECK.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=THYMUS;
RX   MEDLINE; 97429775.
RA   VICARI A.P., FIGUEROA D.J., HEDRICK J.A., FOSTER J.S., SINGH K.P.,
RA   MENON S., COPELAND N.G., GILBERT D.J., JENKINS N.A., BACON K.B.,
RA   ZLOTNIK A.;
RT   "TECK: a novel CC chemokine specifically expressed by thymic
RT   dendritic cells and potentially involved in T cell development.";
RL   IMMUNITY 7:291-301(1997).
CC   -!- FUNCTION: POTENTIALLY INVOLVED IN T CELL DEVELOPMENT. RECOMBINANT
CC       TECK SHOWS CHEMOTACTIC ACTIVITY ON THYMOCYTES, MACROPHAGES, THP-1
CC       CELLS, AND DENDRITICS CELLS BUT IS INACTIVE ON PERIPHERAL BLOOD
CC       LYMPHOCYTES AND NEUTROPHILS.
CC   -!- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED BY THYMIC DENDRITIC
CC       CELLS. HIGH LEVELS IN THYMUS AND SMALL INTESTINE.
CC   -!- SIMILARITY: BELONGS TO THE INTERCRINE BETA FAMILY (SMALL CYTOKINE
CC       C-C) (CHEMOKINE CC).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U86357; G2388629; -.
DR   MGD; MGI:1099448; SCYA25.
DR   PROSITE; PS00472; SMALL_CYTOKINES_CC; FALSE_NEG.
KW   CYTOKINE; CHEMOTAXIS; INFLAMMATORY RESPONSE; SIGNAL.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24    144       CHEMOKINE TECK.
FT   DISULFID     30     58       BY SIMILARITY.
FT   DISULFID     31     73       BY SIMILARITY.
SQ   SEQUENCE   144 AA;  16714 MW;  14D81ED4 CRC32;

  Query Match             43.9%;  Score 312;  DB 1;  Length 144;
  Best Local Similarity   48.6%;  Pred. No. 3e-30;
  Matches   69;  Conservative   28;  Mismatches   33;  Indels   12;  Gaps    4;

Qy       1 MNLWLLACLVAGFLGAWAPAVHAQGVFEDCCLAYHYPIGWAVLRRAWTYRIQEVSGSCNL 60
           |:|||:|||||  |:|||  | |||||  ||||||:|:   | |  |||:|: |: |||||||||
Db       1 MKLWLFACLVACFVGAWMPVVHAQGAFEDCCLGYQHRIKWNVLRHARNYHQQEVSGSCNL 60

Qy      61 PAAIFYLPKRHRKVCGNPKSREVQRAMKLLDARNKVFAKLRHNTQTFQAGPHAVKKLSSG 120
             |  ||:  |:: |||||   :| ||:::| ||:    :| |   :  : ::: :|
Db      61 RAVRFYF--RQKVVCGNPEDMNVKRAIRILTARK----RLVHWKSA--SDSQTERK---- 109

Qy     121 NSKLSSSKFSNPISSSKRNVSL 142
           :|:   ||  || |:| |: :|
Db     109 KSNHMKSKVENPNSTSVRSATL 130


RESULT   3
MI1G_MOUSE
ID   MI1G_MOUSE     STANDARD;      PRT;   122 AA.
AC   P51670;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   MACROPHAGE INFLAMMATORY PROTEIN 1-GAMMA PRECURSOR (MIP-1-GAMMA)
DE   (MACROPHAGE INFLAMMATORY PROTEIN-RELATED PROTEIN-2) (MRP-2) (CCF18).
GN   SCYA9 OR SCYA10 OR MRP2.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96170877.
RA   POLTORAK A.N., BAZZONI F., SMIRNOVA I.I., ALEJOS E., THOMPSON P.,
RA   LUHESHI G., ROTHWELL N., BEUTLER B.;
RT   "MIP-1 gamma: molecular cloning, expression, and biological
RT   activities of a novel CC chemokine that is constitutively secreted in
RT   vivo.";
RL   J. INFLAMM. 45:207-219(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95378690.
RA   YOUN B.-S., JANG I.-K., BROXMEYER H.E., COOPER S., JENKINS N.A.,
RA   GILBERT D.J., COPELAND N.G., ELICK T.A., FRASER M.J. JR., KWON B.S.;
RT   "A novel chemokine, macrophage inflammatory protein-related
RT   protein-2, inhibits colony formation of bone marrow myeloid
RT   progenitors.";
RL   J. IMMUNOL. 155:2661-2667(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96072813.
RA   HARA T., BACON K.B., CHO L.C., YOSHIMURA A., MORIKAWA Y.,
RA   COPELAND N.G., GILBERT D.J., JENKINS N.A., SCHALL T.J., MIYAJIMA A.;
RT   "Molecular cloning and functional characterization of a novel member
RT   of the C-C chemokine family.";
RL   J. IMMUNOL. 155:5352-5358(1995).
CC   -!- FUNCTION: MONOKINE WITH INFLAMMATORY, PYROGENIC AND CHEMOKINETIC
CC       PROPERTIES. CIRCULATES AT HIGH CONCENTRATIONS IN THE BLOOD OF
CC       HEALTHY ANIMALS. BINDING TO A HIGH-AFFINITY RECEPTOR ACTIVATES
CC       CALCIUM RELEASE IN NEUTROPHILS. IT ALSO INHIBITS COLONY FORMATION
CC       OF BONE MARROW MYELOID IMMATURE PROGENITORS.
CC   -!- TISSUE SPECIFICITY: EXPRESSED MAINLY IN THE LIVER, LUNG, AND THE
CC       THYMUS, ALTHOUGH SOME EXPRESSION HAS BEEN DETECTED IN A WIDE
CC       VARIETY OF TISSUES EXCEPT BRAIN.
CC   -!- INDUCTION: BY IL-4 IN THE BONE MARROW MACROPHAGE.
CC   -!- SIMILARITY: BELONGS TO THE INTERCRINE BETA FAMILY (SMALL CYTOKINE
CC       C-C) (CHEMOKINE CC).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U49513; G1234835; -.
DR   EMBL; U15209; G915276; -.
DR   EMBL; U19482; G1101007; -.
DR   MGD; MGI:104533; SCYA9.
DR   PROSITE; PS00472; SMALL_CYTOKINES_CC; 1.
DR   PFAM; PF00048; il8; 1.
DR   HSSP; P13236; 1HUN.
KW   CYTOKINE; CHEMOTAXIS; SIGNAL.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    122       MACROPHAGE INFLAMMATORY PROTEIN 1-GAMMA.
FT   DISULFID     57     80       BY SIMILARITY.
FT   DISULFID     58     96       BY SIMILARITY.
FT   CONFLICT    113    113       Q -> K (IN REF. 1).
SQ   SEQUENCE   122 AA;  13871 MW;  BE281034 CRC32;

  Query Match             13.8%;  Score 98;  DB 1;  Length 122;
  Best Local Similarity   30.9%;  Pred. No. 8.6e-05;
  Matches   34;  Conservative   11;  Mismatches   19;  Indels   46;  Gaps    6;
```

```
Qy      14 LGAWAPAVHA------------QGV--------FE---DCCLAYHYPIGWAVLRRAWTY 49
           || ||   ||             ||:        |:   ||||:|:
Db      17 LGIWAQITHATETKEVQSSLKAQQGLEIEMFHMGFQDSSDCCLSYN------------S 63

Qy      50 RIQ--------EVSGSCNLPAAIFYLPKRHRKVCGNPKSREVQRAMKLLD 91
           |||          ||:|  |: || ::||   ||:||  | ||| :  |:
Db      64 RIQCSRFIGYFPTSGGCTRPGIIF-ISKRGFQVCANPSDRRVQRCIERLE 112


RESULT   4
MIP3_HUMAN
ID   MIP3_HUMAN     STANDARD;      PRT;   121 AA.
AC   P55773;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   MACROPHAGE INFLAMMATORY PROTEIN 3 PRECURSOR (MIP-3).
GN   MIP3.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   LI H., RUBEN S.;
RT   "Macrophage inflammatory protein-3 and -4.";
RL   PATENT NUMBER US5504003, 02-APR-1996.
RN   [2]
RP   DISCUSSION OF SEQUENCE.
RX   MEDLINE; 97275308.
RA   WELLS T.N.C., PEITSCH M.C.;
RT   "The chemokine information source: identification and
RT   characterization of novel chemokines using the WorldWideWeb and
RT   expressed sequence tag databases.";
RL   J. LEUKOC. BIOL. 61:545-550(1997).
CC   -!- FUNCTION: CHEMOTACTIC FACTOR THAT ATTRACTS MONOCYTES, BUT NOT
CC       NEUTROPHILS. THIS PROTEIN CAN BIND HEPARIN.
CC   -!- SIMILARITY: BELONGS TO THE INTERCRINE BETA FAMILY (SMALL CYTOKINE
CC       C-C) (CHEMOKINE CC).
DR   PROSITE; PS00472; SMALL_CYTOKINES_CC; 1.
DR   PFAM; PF00048; il8; 1.
DR   HSSP; P13236; 1HUN.
KW   CYTOKINE; CHEMOTAXIS; HEPARIN-BINDING; SIGNAL.
FT   SIGNAL        1      ?       POTENTIAL.
FT   CHAIN         ?    121       MACROPHAGE INFLAMMATORY PROTEIN 3.
FT   DISULFID     55     79       BY SIMILARITY.
FT   DISULFID     56     95       BY SIMILARITY.
FT   DISULFID     66    106       POTENTIAL.
SQ   SEQUENCE   121 AA;  13998 MW;  A970B450 CRC32;

  Query Match             12.5%;  Score 89;  DB 1;  Length 121;
  Best Local Similarity   26.7%;  Pred. No. 0.001;
  Matches   32;  Conservative   21;  Mismatches   27;  Indels   40;  Gaps    7;

Qy       6 LACLVAGFLGAWAPA---------VHA--------------QGVFEDCCLAYHYPIGWA 41
           |:|| | || :  |:         ||                ::   |||::|
Db       8 LSCLHA-FLLPLVPGPGHKRCRDRVHECQSFHWKIQYFWDRFHATSADCCISY------- 60

Qy      42 VLRRAWTYRIQ----EVSGSCNLPAAIFYLPKRHRKVCGNPKSREVQRAMKL--LDARNK 95
              |:    :     | :: |: |: || | |: |: |:||  ::||  |::  ||:|:|
Db      60 -TPRSIPCSLLESYFETNSECSKPGVIF-LTKKGRRFCANPSDKQVQVCMRMLKLDTRIK 117


RESULT   5
TARC_HUMAN
ID   TARC_HUMAN     STANDARD;      PRT;    94 AA.
AC   Q92583;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   THYMUS AND ACTIVATION-REGULATED CHEMOKINE PRECURSOR (CC CHEMOKINE
DE   TARC).
GN   SCYA17 OR TARC OR A-152E5.3.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 24-32.
RC   TISSUE=PERIPHERAL BLOOD MONOCYTES;
RX   MEDLINE; 96355526.
RA   IMAI T., YOSHIDA T., BABA M., NISHIMURA M., KAKIZAKI M., YOSHIE O.;
RT   "Molecular cloning of a novel T cell-directed CC chemokine expressed
RT   in thymus by signal sequence trap using Epstein-Barr virus vector.";
RL   J. BIOL. CHEM. 271:21514-21521(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   ADAMS M.D., LOFTUS B.J., ZHOU L., CROSBY M., FUHRMANN J., MASON T.M.,
RA   BRANDON R., KIM U.J., KERLAVAGE A.R., VENTER J.C.;
RL   SUBMITTED (JUN-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: CHEMOTACTIC FACTOR FOR T LYMPHOCYTES BUT NOT MONOCYTES
CC       OR GRANULOCYTES. MAY PLAY A ROLE IN T CELL DEVELOPEMENT IN THYMUS
CC       AND IN TRAFFICKING AND ACTIVATION OF MATURE T CELLS.
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN THYMUS AND AT LOW
CC       LEVELS IN THE LUNG, COLON AND SMALL INTESTINE.
CC   -!- INDUCTION: BY PHYTOHEMAGGLUTININ (PHA) IN THE PERIPHERAL BLOOD
CC       MONONUCLEAR CELLS AND BY CYTOKINES IN MONOCYTES.
CC   -!- SIMILARITY: BELONGS TO THE INTERCRINE BETA FAMILY (SMALL CYTOKINE
CC       C-C) (CHEMOKINE CC).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D43767; G1536879; -.
DR   EMBL; AC004382; G3252822; -.
DR   MIM; 601520; -.
DR   PROSITE; PS00472; SMALL_CYTOKINES_CC; 1.
DR   PFAM; PF00048; il8; 1.
KW   CYTOKINE; CHEMOTAXIS; INFLAMMATORY RESPONSE; SIGNAL.
FT   SIGNAL        1     23
FT   CHAIN        24     94       THYMUS AND ACTIVATION-REGULATED
FT                                CHEMOKINE.
FT   DISULFID     33     57       BY SIMILARITY.
FT   DISULFID     34     73       BY SIMILARITY.
SQ   SEQUENCE   94 AA;  10507 MW;  4E4FC818 CRC32;

  Query Match             11.5%;  Score 81.5;  DB 1;  Length 94;
  Best Local Similarity   31.6%;  Pred. No. 0.0059;
  Matches   31;  Conservative   23;  Mismatches   31;  Indels   13;  Gaps    7;

Qy       3 LWLLACLVAGFLGAWAPAVHA---QGVFEDCCLAYHYPIGWAVLR--RAWTYRIQEVSGS 57
           | :|| ||: :|||    :||     |  :||| |    |   ||  ::| |:  |
Db       4 LKMLA-LVTLLLGASLQHIHAARGTNVGRECCLEYF--KGAIPLRKLKTW-YQTSE---D 56

Qy      58 CNLPAAIFYLPKRHRKVCGNPKSREVQRAMKLLDARNK 95
           |:   || ::  : | :|::|::: |  |:| |:: ::
Db      57 CS-RDAIVFVTVQGRAICSDPNNKRVKNAVKYLQSLER 93


RESULT   6
C10_MOUSE
ID   C10_MOUSE      STANDARD;      PRT;   116 AA.
AC   P27784;
DT   01-AUG-1992 (REL. 23, CREATED)
DT   01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   C10 PROTEIN PRECURSOR.
GN   SCYA6 OR C10.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
```

```
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CBA/J; TISSUE=BONE MARROW;
RX   MEDLINE; 91370083.
RA   ORLOFSKY A., BERGER M.S., PRYSTOWSKY M.B.;
RT   "Novel expression pattern of a new member of the MIP-1 family of
RT   cytokine-like genes.";
RL   CELL REGUL. 2:403-412(1991).
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN MYELOPOIETIC BONE MARROW
CC       CULTURES STIMULATED BY GM-CSF.
CC   -!- INDUCTION: ASSOCIATED WITH STIMULI THAT PROMOTE MYELOID
CC       DIFFERENTIATION.
CC   -!- SIMILARITY: BELONGS TO THE INTERCRINE BETA FAMILY (SMALL CYTOKINE
CC       C-C) (CHEMOKINE CC).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M58004; G192244; -.
DR   MGD; MGI:98263; SCYA6.
DR   PROSITE; PS00472; SMALL_CYTOKINES_CC; 1.
DR   PFAM; PF00048; il8; 1.
DR   HSSP; P13236; 1HUN.
KW   CYTOKINE; CHEMOTAXIS; SIGNAL.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    116       C10 PROTEIN.
FT   DISULFID     50     73       BY SIMILARITY.
FT   DISULFID     51     89       BY SIMILARITY.
SQ   SEQUENCE   116 AA;  12984 MW;  A745B62D CRC32;
```

```
  Query Match             11.4%;  Score 81;  DB 1;  Length 116;
  Best Local Similarity   26.7%;  Pred. No. 0.0087;
  Matches   23;  Conservative   13;  Mismatches   20;  Indels   30;  Gaps    4;

Qy      18 APAVHA--QGVFEDCCLA----------YHYPIGWAVLRRAWTYRIQEVSGSCNLPAAI 64
           :| :|   |    |||::          | :|                ||:|  |: |
Db      36 PPIIHQGFQDTSSDCCFSYATQIPCKRFIYYFP---------------TSGGCIKPGII 79

Qy      65 FYLPKRHRKVCGNPKSREVQRAMKLL 90
           | :::|   ||::|  | |||  :  |
Db      80 F-ISRRGTQVCADPSDRRVQRCLSTL 104
```

```
RESULT   7
MI1A_RAT
ID   MI1A_RAT       STANDARD;      PRT;    92 AA.
AC   P50229;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   MACROPHAGE INFLAMMATORY PROTEIN 1-ALPHA PRECURSOR (MIP-1-ALPHA).
GN   SCYA3 OR MIP1A.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CD-1; TISSUE=LUNG;
RX   MEDLINE; 95298037.
RA   SHI M.M., GODLESKI J.J., PAULAUSKIS J.D.;
RT   "Molecular cloning and posttranscriptional regulation of macrophage
RT   inflammatory protein-1 alpha in alveolar macrophages.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 211:289-295(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LONG EVANS; TISSUE=LUNG;
RX   MEDLINE; 95238980.
RA   SHANLEY T.P., SCHMAL H., FRIEDL H.P., JONES M.L., WARD P.A.;
RT   "Role of macrophage inflammatory protein-1 alpha (MIP-1 alpha) in
RT   acute lung injury in rats.";
RL   J. IMMUNOL. 154:4793-4802(1995).
RN   [3]
RP   SEQUENCE OF 24-57.
RC   STRAIN=WISTAR;
RX   MEDLINE; 96183056.
RA   NAKAGAWA H., SHIOTA S., TAKANO K., SHIBATA F., KATO H.;
RT   "Cytokine-induced neutrophil chemoattractant (CINC)-2 alpha, a novel
RT   member of rat GRO/CINCs, is a predominant chemokine produced by
RT   lipopolysaccharide-stimulated rat macrophages in culture.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 220:945-948(1996).
CC   -!- FUNCTION: MONOKINE WITH INFLAMMATORY AND CHEMOKINETIC PROPERTIES.
CC       HAS CHEMOTACTIC ACTIVITY FOR MONOCYTES, NEUTROPHILS, EOSINOPHILS,
CC       BASOPHILS, AND LYMPHOCYTES. REQUIRED FOR LUNG TNF-ALPHA
CC       PRODUCTION, NEUTROPHIL RECRUITMENT AND SUBSEQUENT LUNG INJURY AND
CC       MAY FUNCTION AS AN AUTOCRINE MEDIATOR FOR THE MACROPHAGE
CC       PRODUCTION OF TNF-ALPHA WHICH IN TURN UP-REGULATES VASCULAR
CC       ADHESION MOLECULES REQUIRED FOR NEUTROPHIL INFLUX. THIS PROTEIN
CC       BINDS HEPARIN.
CC   -!- INDUCTION: BY LIPOPOLYSACCHARIDE (LPS).
CC   -!- SIMILARITY: BELONGS TO THE INTERCRINE BETA FAMILY (SMALL CYTOKINE
CC       C-C) (CHEMOKINE CC).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U22414; G790633; -.
DR   EMBL; U06435; G459150; -.
DR   PROSITE; PS00472; SMALL_CYTOKINES_CC; 1.
DR   PFAM; PF00048; il8; 1.
DR   HSSP; P13236; 1HUN.
KW   CYTOKINE; CHEMOTAXIS; INFLAMMATORY RESPONSE; SIGNAL; HEPARIN-BINDING.
FT   SIGNAL        1     23
FT   CHAIN        24     92       MACROPHAGE INFLAMMATORY PROTEIN 1-ALPHA.
FT   DISULFID     34     57       BY SIMILARITY.
FT   DISULFID     35     73       BY SIMILARITY.
FT   CONFLICT      6      6       A -> T (IN REF. 2).
FT   CONFLICT     57     57       C -> W (IN REF. 2 AND 3).
SQ   SEQUENCE   92 AA;  10335 MW;  F48CF89F CRC32;
```

```
  Query Match             11.3%;  Score 80;  DB 1;  Length 92;
  Best Local Similarity   27.2%;  Pred. No. 0.0087;
  Matches   22;  Conservative   19;  Mismatches   20;  Indels   20;  Gaps    6;

Qy       1 MNLWLLACLVAGFLGAWAPAVHAQGVFEDCCLAYHYPIGWAVLRRAWTYRIQ--EVSGSC 58
           | ||        | :: ::   :      ||::|    |: : |:   :  :  | |: |
Db      15 MALW-----NEVFSAPYGADTPT-----ACCFSY----GRQIPRK---FIADYFETSSLC 57

Qy      59 NLPAAIFYLPKRHRKVCGNPK 79
           : |: || | ||:| :|::||
Db      58 SQPGVIF-LTKRNRQICADPK 77
```

```
RESULT   8
MIP4_HUMAN
ID   MIP4_HUMAN     STANDARD;      PRT;    89 AA.
AC   P55774;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   MACROPHAGE INFLAMMATORY PROTEIN 4 PRECURSOR (MIP-4) (PULMONARY AND
DE   ACTIVATION-REGULATED CHEMOKINE) (CC CHEMOKINE PARC) (ALTERNATIVE
DE   ACTIVATED MACROPHAGE ASSOCIATED CC CHEMOKINE 1) (AMAC-1).
GN   SCYA18 OR MIP4.
```

```
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   LI H., RUBEN S.;
RT   "Macrophage inflammatory protein-3 and -4.";
RL   PATENT NUMBER US5504003, 02-APR-1996.
RN   [2]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=AORTA, AND LUNG;
RX   MEDLINE; 97376836.
RA   HIESHIMA K., IMAI T., BABA M., SHOUDAI K., ISHIZUKA K.,
RA   NAKAGAWA T., TSURUTA J., TAKEYA M., SAKAKI Y., TAKATSUKI K.,
RA   MIURA R., OPDENAKKER G., VAN DAMME J., YOSHIE O., NOMIYAMA H.;
RT   "A novel human CC chemokine PARC that is most homologous to
RT   macrophage-inflammatory protein-1 alpha/LD78 alpha and chemotactic
RT   for T lymphocytes, but not for monocytes.";
RL   J. IMMUNOL. 159:1140-1149(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   KODELJA V., MUELLER C., POLITZ O., HAKIY N., ORFANOS C.E., GOERDT S.;
RL   SUBMITTED (AUG-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [4]
RP   DISCUSSION OF SEQUENCE.
RX   MEDLINE; 97275308.
RA   WELLS T.N.C., PEITSCH M.C.;
RT   "The chemokine information source: identification and
RT   characterization of novel chemokines using the WorldWideWeb and
RT   expressed sequence tag databases.";
RL   J. LEUKOC. BIOL. 61:545-550(1997).
CC   -!- FUNCTION: CHEMOTACTIC FACTOR THAT ATTRACTS LYMPHOCYTES BUT NOT
CC       MONOCYTES OR GRANULOCYTES. MAY BE INVOLVED IN B CELL MIGRATION
CC       INTO B CELL FOLLICLES IN LYMPH NODES.
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN THE LUNG. A LOWER
CC       LEVEL EXPRESSION IS SEEN IN LYMPHOID TISSUES SUCH AS LYMPH NODES,
CC       THYMUS AND APPENDIX.
CC   -!- INDUCTION: BY LIPOPOLYSACCHARIDE (LPS).
CC   -!- SIMILARITY: BELONGS TO THE INTERCRINE BETA FAMILY (SMALL CYTOKINE
CC       C-C) (CHEMOKINE CC).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB000221; D1022520; -.
DR   EMBL; Y13710; E321838; -.
DR   PROSITE; PS00472; SMALL_CYTOKINES_CC; 1.
DR   PFAM; PF00048; il8; 1.
DR   HSSP; P13236; 1HUN.
KW   CYTOKINE; CHEMOTAXIS; INFLAMMATORY RESPONSE; SIGNAL.
FT   SIGNAL        1     20
FT   CHAIN        21     89       MACROPHAGE INFLAMMATORY PROTEIN 4.
FT   DISULFID     30     54       BY SIMILARITY.
FT   DISULFID     31     70       BY SIMILARITY.
SQ   SEQUENCE   89 AA;  9849 MW;  052AA3DC CRC32;


  Query Match             11.1%;  Score 79;  DB 1;  Length 89;
  Best Local Similarity   30.8%;  Pred. No. 0.011;
  Matches   20;  Conservative   16;  Mismatches   17;  Indels   12;  Gaps    5;

Qy      30 CCLAYHYPIGWAVLRRAWTYRIQ--EVSGSCNLPAAIFYLPKRHRKVCGNPKSREVQR-- 86
           ||| |    :| : ::   : ::  | |  |  |: |: | || | :|::|: : ||:
Db      30 CCLVY---TSWQIPQK---FIVDYSETSPQCPKPGVIL-LTKRGRQICADPNKKWVQKYI 82

Qy      86 -AMKL 89
            :||
Db      83 SDLKL 87


RESULT   9
EOTA_HUMAN
ID   EOTA_HUMAN     STANDARD;      PRT;    97 AA.
AC   P51671; P50877; Q92490; Q92491;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   EOTAXIN PRECURSOR (EOSINOPHIL CHEMOTACTIC PROTEIN).
GN   SCYA11.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96181758.
RA   GARCIA-ZEPEDA E.A., ROTHENBERG M.E., OWNBEY T.R., LEDER P.,
RA   LUSTER A.D.;
RT   "Human eotaxin is a specific chemoattractant for eosinophil cells and
RT   provides a new mechanism to explain tissue eosinophilia.";
RL   NAT. MED. 2:449-456(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96189937.
RA   PONATH P.D., QIN S., RINGLER D.J., CLARK-LEWIS I., WANG J., KASSAM N.,
RA   SMITH H., SHI X., GONZALO J.A., NEWMAN W., GUTIERREZ-RAMOS J.C.,
RA   MACKAY C.R.;
RT   "Cloning of the human eosinophil chemoattractant, eotaxin.
RT   Expression, receptor binding, and functional properties suggest a
RT   mechanism for the selective recruitment of eosinophils.";
RL   J. CLIN. INVEST. 97:604-612(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SMALL INTESTINE;
RX   MEDLINE; 96205964.
RA   KITAURA M., NAKAJIMA T., IMAI T., HARADA S., COMBADIERE C.,
RA   TIFFANY H.L., MURPHY P.M., YOSHIE O.;
RT   "Molecular cloning of human eotaxin, an eosinophil-selective CC
RT   chemokine, and identification of a specific eosinophil eotaxin
RT   receptor, CC chemokine receptor 3.";
RL   J. BIOL. CHEM. 271:7725-7730(1996).
RN   [4]
RP   SEQUENCE FROM N.A., SEQUENCE OF 60-65 AND 75-88, AND VARIANTS.
RC   TISSUE=FORESKIN;
RX   MEDLINE; 96374440.
RA   BARTELS J., SCHLUETER C., RICHTER E., NOSO N., KULKE R.,
RA   CHRISTOPHERS E., SCHROEDER J.M.;
RT   "Human dermal fibroblasts express eotaxin: molecular cloning, mRNA
RT   expression, and identification of eotaxin sequence variants.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 225:1045-1051(1996).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE; 97312708.
RA   GARCIA-ZEPEDA E.A., ROTHENBERG M.E., WEREMOWICZ S., SARAFI M.N.,
RA   MORTON C.C., LUSTER A.D.;
RT   "Genomic organization, complete sequence, and chromosomal location of
RT   the gene for human eotaxin (SCYA11), an eosinophil-specific CC
RT   chemokine.";
RL   GENOMICS 41:471-476(1997).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LUNG;
RX   MEDLINE; 97445071.
RA   HEIN H., SCHLUETER C., KULKE R., CHRISTOPHERS E., SCHROEDER J.M.,
RA   BARTELS J.;
RT   "Genomic organization, sequence, and transcriptional regulation of
RT   the human eotaxin gene.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 237:537-542(1997).
RN   [7]
RP   STRUCTURE BY NMR.
RX   MEDLINE; 98380469.
```

```
RA   CRUMP M.P., RAJARATHNAM K., KIM K.S., CLARK-LEWIS I., SYKES B.D.;
RT   "Solution structure of eotaxin, a chemokine that selectively recruits
RT   eosinophils in allergic inflammation.";
RL   BIOCHEMISTRY 37:11670-11678(1998).
CC   -!- FUNCTION: IN RESPONSE TO THE PRESENCE OF ALLERGENS, THIS PROTEIN
CC       DIRECTLY PROMOTES THE ACCUMULATION OF EOSINOPHILS, A PROMINENT
CC       FEATURE OF ALLERGIC INFLAMMATORY REACTIONS.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- PTM: O-GLYCOSYLATED (PROBABLE).
CC   -!- INDUCTION: BY TNF-ALPHA, IL-1-ALPHA AND INTERFERON GAMMA.
CC   -!- SIMILARITY: BELONGS TO THE INTERCRINE BETA FAMILY (SMALL CYTOKINE
CC       C-C) (CHEMOKINE CC).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U46573; G1280141; -.
DR   EMBL; U34780; G1185440; -.
DR   EMBL; D49372; G1552241; -.
DR   EMBL; Z69291; E221070; -.
DR   EMBL; Z75668; E251275; -.
DR   EMBL; Z75669; E251258; -.
DR   EMBL; U46572; G2088509; -.
DR   EMBL; Z92709; E329504; -.
DR   MIM; 601156; -.
DR   PROSITE; PS00472; SMALL_CYTOKINES_CC; 1.
DR   PFAM; PF00048; il8; 1.
DR   PDB; 2EOT; 11-NOV-98.
KW   EOSINOPHIL; CYTOKINE; CHEMOTAXIS; GLYCOPROTEIN; SIGNAL;
KW   INFLAMMATORY RESPONSE; POLYMORPHISM; 3D-STRUCTURE.
FT   SIGNAL       1     23
FT   CHAIN       24     97       EOTAXIN.
FT   DISULFID    32     57
FT   DISULFID    33     73
FT   VARIANT      7      7       L -> P (IN CLONE 34).
FT   VARIANT     23     23       A -> T (IN CLONE 53).
FT   VARIANT     51     51       R -> S (IN CLONE 34).
FT   VARIANT     79     79       K -> R (IN CLONE 53).
SQ   SEQUENCE   97 AA;  10732 MW;  6C0F3D98 CRC32;
```

```
  Query Match             11.1%;  Score 78.5;  DB 1;  Length 97;
  Best Local Similarity   33.7%;  Pred. No. 0.014;
  Matches   31;  Conservative   21;  Mismatches   33;  Indels    7;  Gaps    5;

Qy       3 LWLLACLVAGFLGAWAPAVHAQGVFEDCCLAYHYPIGWAVLRRAWTYRIQEVSGSCNLPA 62
           |||| |:|: ::: : |  | :|   ||  :: :     |:|  :|| :  || |   |
Db       8 LWLL--LIAAAFSPQGLAGPA-SVPTTCC--FNLANRKIPLQRLESYR-RITSGKCPQKA 61

Qy      63 AIFYLPKRHRKVCGNPKSREVQRAMKLLDARN 94
             ||   |  : :|::|| : || :|| || ::
Db      62 VIF-KTKLAKDICADPKKKWVQDSMKYLDQKS 92
```

```
RESULT  10
MIP5_HUMAN
ID   MIP5_HUMAN     STANDARD;      PRT;   113 AA.
AC   Q16663;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   MACROPHAGE INFLAMMATORY PROTEIN 5 PRECURSOR (MIP-5) (CHEMOKINE CC-2)
DE   (HCC-2) (NCC-3) (MIP-1 DELTA).
GN   SCYA15 OR MIP5 OR NCC3.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   PARDIGOL A., MAEGERT H.-J., ZUCHT H.D., FORSSMANN W.-G.,
RA   SCHULZ-KNAPPE P.;
RL   SUBMITTED (MAY-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   WANG W.;
RL   SUBMITTED (OCT-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE OF 12-113 FROM N.A.
RA   COULIN F., POWER C.A., ALOUANI S., PEITSCH M.C., SCHROEDER J.-M.,
RA   MOSHIZUKI M., CLARK-LEWIS I., WELLS T.N.C.;
RL   SUBMITTED (JAN-1997) TO THE SWISS-PROT DATA BANK.
RN   [4]
RP   DISCUSSION OF SEQUENCE.
RX   MEDLINE; 97275308.
RA   WELLS T.N.C., PEITSCH M.C.;
RT   "The chemokine information source: identification and
RT   characterization of novel chemokines using the WorldWideWeb and
RT   expressed sequence tag databases.";
RL   J. LEUKOC. BIOL. 61:545-550(1997).
CC   -!- FUNCTION: CHEMOTACTIC FACTOR THAT ATTRACTS MONOCYTES, BUT NOT
CC       NEUTROPHILS.
CC   -!- SIMILARITY: BELONGS TO THE INTERCRINE BETA FAMILY (SMALL CYTOKINE
CC       C-C) (CHEMOKINE CC).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF031587; G2739164; -.
DR   EMBL; Z70293; E233857; -.
DR   EMBL; Z70292; E233855; -.
DR   MIM; 601393; -.
DR   PROSITE; PS00472; SMALL_CYTOKINES_CC; 1.
DR   PFAM; PF00048; il8; 1.
DR   HSSP; P13501; 1HRJ.
KW   CYTOKINE; CHEMOTAXIS; HEPARIN-BINDING; SIGNAL.
FT   SIGNAL       1     21       POTENTIAL.
FT   CHAIN       22    113       MACROPHAGE INFLAMMATORY PROTEIN 5.
FT   DISULFID    53     77       BY SIMILARITY.
FT   DISULFID    54     93       BY SIMILARITY.
FT   DISULFID    64    104       POTENTIAL.
FT   CONFLICT    24     24       I -> T (IN REF. 1).
SQ   SEQUENCE   113 AA;  12248 MW;  A234178D CRC32;
```

```
  Query Match             10.8%;  Score 76.5;  DB 1;  Length 113;
  Best Local Similarity   35.4%;  Pred. No. 0.029;
  Matches   23;  Conservative   13;  Mismatches   18;  Indels   11;  Gaps    4;

Qy      29 DCCLAYHYPIGWAV---LRRAWTYRIQEVSGSCNLPAAIFYLPKRHRKVCGNPKSREVQR 85
           ||| :|   |: ::   | ::      | |: |: |: || | |: | ||::| :  ||
Db      52 DCCTSY---ISQSIPCSLMKSYF----ETSSECSKPGVIF-LTKKGRQVCAKPSGPGVQD 103

Qy      86 AMKLL 90
             || |
Db     104 CMKKL 108
```

```
RESULT  11
SISD_HUMAN
ID   SISD_HUMAN     STANDARD;      PRT;     91 AA.
AC   P13501;
DT   01-JAN-1990 (REL. 13, CREATED)
DT   01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   T-CELL SPECIFIC RANTES PROTEIN PRECURSOR (SIS-DELTA) (SMALL INDUCIBLE
```

```
DE   CYTOKINE A5).
GN   SCYA5.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88285659.
RA   SCHALL T.J., JONGSTRA J., DYER B.J., JORGENSEN J., CLAYBERGER C.,
RA   DAVIS M.M., KRENSKY A.M.;
RT   "A human T cell-specific molecule is a member of a new gene family.";
RL   J. IMMUNOL. 141:1018-1025(1988).
RN   [2]
RP   STRUCTURE BY NMR.
RX   MEDLINE; 95352612.
RA   CHUNG C.-W., COOKE R.M., PROUDFOOT A.E.I., WELLS T.N.C.;
RT   "The three-dimensional solution structure of RANTES.";
RL   BIOCHEMISTRY 34:9307-9314(1995).
RN   [3]
RP   STRUCTURE BY NMR.
RX   MEDLINE; 95244456.
RA   SKELTON N.J., ASPIRAS F., OGEZ J., SCHALL T.J.;
RT   "Proton NMR assignments and solution conformation of RANTES, a
RT   chemokine of the C-C type.";
RL   BIOCHEMISTRY 34:5329-5342(1995).
CC   -!- FUNCTION: CHEMOATTRACTANT FOR BLOOD MONOCYTES, MEMORY T HELPER
CC       CELLS AND EOSINOPHILS. CAUSES THE RELEASE OF HISTAMINE FROM
CC       BASOPHILS AND ACTIVATES EOSINOPHILS.
CC   -!- TISSUE SPECIFICITY: T-CELL AND MACROPHAGE SPECIFIC.
CC   -!- INDUCTION: BY MITOGENS.
CC   -!- SIMILARITY: BELONGS TO THE INTERCRINE BETA FAMILY (SMALL CYTOKINE
CC       C-C) (CHEMOKINE CC).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M21121; G339421; -.
DR   PIR; A28815; A28815.
DR   PDB; 1HRJ; 14-OCT-96.
DR   PDB; 1RTN; 03-JUN-95.
DR   PDB; 1RTO; 03-JUN-95.
DR   MIM; 187011; -.
DR   PROSITE; PS00472; SMALL_CYTOKINES_CC; 1.
DR   PFAM; PF00048; il8; 1.
KW   CYTOKINE; CHEMOTAXIS; T-CELL; SIGNAL; INFLAMMATORY RESPONSE;
KW   3D-STRUCTURE.
FT   SIGNAL        1     23
FT   CHAIN        24     91       T-CELL SPECIFIC RANTES PROTEIN.
FT   DISULFID     33     57
FT   DISULFID     34     73
SQ   SEQUENCE   91 AA;  10075 MW;  6BDFE23E CRC32;


  Query Match             10.8%;  Score 76.5;  DB 1;  Length 91;
  Best Local Similarity   35.4%;  Pred. No. 0.023;
  Matches   23;  Conservative   17;  Mismatches   14;  Indels   11;  Gaps    5;

Qy      30 CCLAYHYPIGWAVLRRAWTYRIQE---VSGSCNLPAAIFYLPKRHRKVCGNPKSREVQRA 86
           ||:||   |::: | ||   :| |    || |: || :| : :::| ||:||  : |:
Db      33 CCFAY---IARP-LPRA---HIKEYFYTSGKCSNPAVVF-VTRKNRQVCANPEKKWVREY 84

Qy      87 MKLLD 91
           :: |:
Db      85 INSLE 89


RESULT  12
MI3A_HUMAN
ID   MI3A_HUMAN     STANDARD;      PRT;    96 AA.
AC   P78556;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   MACROPHAGE INFLAMMATORY PROTEIN 3 ALPHA PRECURSOR (MIP-3-ALPHA) (LIVER
DE   AND ACTIVATION-REGULATED CHEMOKINE) (CC CHEMOKINE LARC).
GN   SCYA20 OR MIP3A OR LARC.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97166046.
RA   ROSSI D.L., VICARI A.P., FRANZ-BACON K., MCCLANAHAN T.K., ZLOTNIK A.;
RT   "Identification through bioinformatics of two new macrophage
RT   proinflammatory human chemokines: MIP-3alpha and MIP-3beta.";
RL   J. IMMUNOL. 158:1033-1036(1997).
RN   [2]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=LIVER;
RX   MEDLINE; 97190319.
RA   HIESHIMA K., IMAI T., OPDENAKKER G., VAN DAMME J., KUSUDA J.,
RA   TEI H., SAKAKI Y., TAKATSUKI K., MIURA R., YOSHIE O., NOMIYAMA H.;
RT   "Molecular cloning of a novel human CC chemokine liver and
RT   activation-regulated chemokine (LARC) expressed in liver.
RT   Chemotactic activity for lymphocytes and gene localization on
RT   chromosome 2.";
RL   J. BIOL. CHEM. 272:5846-5853(1997).
CC   -!- FUNCTION: CHEMOTACTIC FACTOR THAT ATTRACTS LYMPHOCYTES AND,
CC       SLIGHTLY, NEUTROPHILS, BUT NOT MONOCYTES.
CC   -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN THE LIVER, LYMPH
CC       NODES, APPENDIX, PBL, AND FETAL LUNG. LOW LEVELS SEEN IN THYMUS,
CC       PROSTATE, TESTIS, SMALL INTESTINE AND COLON.
CC   -!- INDUCTION: BY LIPOPOLYSACCHARIDE (LPS) AND GAMMA-IFN, REPRESSED BY
CC       IL-10.
CC   -!- SIMILARITY: BELONGS TO THE INTERCRINE BETA FAMILY (SMALL CYTOKINE
CC       C-C) (CHEMOKINE CC).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U77035; G1790925; -.
DR   EMBL; D86955; G1871139; -.
DR   PROSITE; PS00472; SMALL_CYTOKINES_CC; 1.
DR   PFAM; PF00048; il8; 1.
DR   HSSP; P13236; 1HUN.
KW   CYTOKINE; CHEMOTAXIS; INFLAMMATORY RESPONSE; SIGNAL.
FT   SIGNAL        1     26
FT   CHAIN        27     96       MACROPHAGE INFLAMMATORY PROTEIN 3 ALPHA.
FT   DISULFID     32     58       BY SIMILARITY.
FT   DISULFID     33     74       BY SIMILARITY.
SQ   SEQUENCE   96 AA;  10762 MW;  DBAC0235 CRC32;


  Query Match             10.7%;  Score 76;  DB 1;  Length 96;
  Best Local Similarity   28.6%;  Pred. No. 0.028;
  Matches   28;  Conservative   17;  Mismatches   35;  Indels   18;  Gaps    4;

Qy       5 LLACLVAGFLGAWAPAVHAQGVFEDCCLAY-------HYPIGWAVLRRAWTYRIQEVSGS 57
           ||| |:: :|        | : | ||||:|        : :|         :  |  : :
Db       8 LLAALMSVLLLHLCGESEAASNF-DCCLGYTDRILHPKFIVG---------FTRQLANEG 57

Qy      58 CNLPAAIFYLPKRHRKVCGNPKSREVQRAMKLLDARNK 95
           |:: | ||   |:   ||:|||   |   ::||  : |
Db      58 CDINAIIFH-TKKKLSVCANPKQTWVKYIVRLLSKKVK 94
```

```
RESULT  13
MI3B_HUMAN
ID   MI3B_HUMAN     STANDARD;      PRT;    98 AA.
AC   Q99731; O00736; O00697;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   MACROPHAGE INFLAMMATORY PROTEIN 3 BETA PRECURSOR (MIP-3-BETA) (EBI1-
DE   LIGAND CHEMOKINE) (ELC) (BETA CHEMOKINE EXODUS-3).
GN   SCYA19 OR MIP3B OR ELC.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97166046.
RA   ROSSI D.L., VICARI A.P., FRANZ-BACON K., MCCLANAHAN T.K., ZLOTNIK A.;
RT   "Identification through bioinformatics of two new macrophage
RT   proinflammatory human chemokines: MIP-3alpha and MIP-3beta.";
RL   J. IMMUNOL. 158:1033-1036(1997).
RN   [2]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=FETAL LUNG;
RX   MEDLINE; 97298088.
RA   YOSHIDA R., IMAI T., HIESHIMA K., KUSUDA J., BABA M., KITAURA M.,
RA   NISHIMURA M., KAKIZAKI M., NOMIYAMA H., YOSHIE O.;
RT   "Molecular cloning of a novel human CC chemokine EBI1-ligand
RT   chemokine that is a specific functional ligand for EBI1, CCR7.";
RL   J. BIOL. CHEM. 272:13803-13809(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   HROMAS R.A., GRAY P., KLEMSZ M., FIFE K., BROXMEYER H.;
RT   "DCCL chemokines represent a novel beta chemokine subfamily.";
RL   SUBMITTED (JUN-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: MAY PLAY A ROLE NOT ONLY IN INFLAMMATORY AND
CC       IMMUNOLOGICAL RESPONSES BUT ALSO IN NORMAL LYMPHOCYTE
CC       RECIRCULATION AND HOMING.
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN THE LYMPH NODES,
CC       THYMUS AND APPENDIX. INTERMEDIATE LEVELS SEEN IN COLON AND
CC       TRACHEA, WHILE LOW LEVELS FOUND IN SPLEEN, SMALL INTESTINE, LUNG,
CC       KIDNEY AND STOMACH.
CC   -!- SIMILARITY: BELONGS TO THE INTERCRINE BETA FAMILY (SMALL CYTOKINE
CC       C-C) (CHEMOKINE CC).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U77180; G1791003; -.
DR   EMBL; AB000887; D1021215; -.
DR   EMBL; U88321; G3243080; -.
DR   MIM; 602227; -.
DR   PROSITE; PS00472; SMALL_CYTOKINES_CC; 1.
DR   PFAM; PF00048; il8; 1.
DR   HSSP; P13236; 1HUN.
KW   CYTOKINE; CHEMOTAXIS; INFLAMMATORY RESPONSE; SIGNAL.
FT   SIGNAL        1     21
FT   CHAIN        22     98       MACROPHAGE INFLAMMATORY PROTEIN 3 BETA.
FT   DISULFID     29     55       BY SIMILARITY.
FT   DISULFID     30     71       BY SIMILARITY.
SQ   SEQUENCE   98 AA;  10993 MW;  897133AC CRC32;

  Query Match             10.4%;  Score 73.5;  DB 1;  Length 98;
  Best Local Similarity   22.7%;  Pred. No. 0.056;
  Matches   25;  Conservative   21;  Mismatches   29;  Indels   35;  Gaps    5;

Qy       9 LVAGFLGAW---APAVHAQGVFEDCCLA------------YHYPIGWAVLRRAWTYRIQ 52
             |  ::|  |   ||:: :    |||||:             :|| :

Db       5 LALSLLVLWTSPAPTLSGTNDAEDCCLSVTQKPIPGYIVRNFHYLL-------------- 51

Qy      53 EVSGSCNLPAAIFYLPKRHRKVCGNPKSREVQRAMKLLDARNKVFAKLRH 102
           :  :| :|| :|    | | :|: |    |:| :  |:   :  ||:::
Db      51 -IKDGCRVPAVVF-TTLRGRQLCAPPDQPWVERIIQRLQ---RTSAKMKR 95


RESULT  14
MI1A_MOUSE
ID   MI1A_MOUSE     STANDARD;      PRT;    92 AA.
AC   P10855; P14096;
DT   01-JUL-1989 (REL. 11, CREATED)
DT   01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   MACROPHAGE INFLAMMATORY PROTEIN 1-ALPHA PRECURSOR (MIP-1-ALPHA) (TY-5)
DE   (SIS-ALPHA) (HEPARIN-BINDING CHEMOTAXIS PROTEIN) (L2G25B).
GN   SCYA3 OR MIP1A.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88258380.
RA   DAVATELIS G., TEKAMP-OLSON P., WOLPE S.D., HERMSEN K., LUEDKE C.,
RA   GALLEGOS C., COIT D., MERRYWEATHER J., CERAMI A.;
RT   "Cloning and characterization of a cDNA for murine macrophage
RT   inflammatory protein (MIP), a novel monokine with inflammatory and
RT   chemokinetic properties.";
RL   J. EXP. MED. 167:1939-1944(1988).
RN   [2]
RP   REVISIONS.
RA   DAVATELIS G., TEKAMP-OLSON P., WOLPE S.D., HERMSEN K., LUEDKE C.,
RA   GALLEGOS C., COIT D., MERRYWEATHER J., CERAMI A.;
RL   J. EXP. MED. 170:2189-2189(1989).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89093958.
RA   BROWN K.D., ZURAWSKI S.M., MOSMANN T.R., ZURAWSKI G.;
RT   "A family of small inducible proteins secreted by leukocytes are
RT   members of a new superfamily that includes leukocyte and
RT   fibroblast-derived inflammatory agents, growth factors, and
RT   indicators of various activation processes.";
RL   J. IMMUNOL. 142:679-687(1989).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DBA/2J;
RX   MEDLINE; 91016858.
RA   GROVE M., LOWE S., GRAHAM G., PRAGNELL I., PLUMB M.;
RT   "Sequence of the murine haemopoietic stem cell inhibitor/macrophage
RT   inflammatory protein 1 alpha gene.";
RL   NUCLEIC ACIDS RES. 18:5561-5561(1990).
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89184547.
RA   KWON B.S., WEISSMAN S.M.;
RT   "cDNA sequences of two inducible T-cell genes.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 86:1963-1967(1989).
RN   [6]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91237116.
RA   WIDMER U., YANG Z., VAN DEVENTER S., MANOGUE K.R., SHERRY B.,
RA   CERAMI A.;
RT   "Genomic structure of murine macrophage inflammatory protein-1 alpha
RT   and conservation of potential regulatory sequences with a human
RT   homolog, LD78.";
RL   J. IMMUNOL. 146:4031-4040(1991).
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/CJ, DBA/2J, NOD/LTJ, SJL/J, AND B10.S/J; TISSUE=SPLEEN;
RA   MA R.Z., TEUSCHER C.;
RL   SUBMITTED (MAY-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [8]
RP   SEQUENCE OF 24-42.
```

```
RX   MEDLINE; 88154745.
RA   WOLPE S.D., DAVATELIS G., SHERRY B., BEUTLER B., HESSE D.G.,
RA   NGUYEN H.T., MOLDAWER L.L., NATHAN C.F., LOWRY S.F., CERAMI A.;
RT   "Macrophages secrete a novel heparin-binding protein with
RT   inflammatory and neutrophil chemokinetic properties.";
RL   J. EXP. MED. 167:570-581(1988).
CC   -!- FUNCTION: MONOKINE WITH INFLAMMATORY, PYROGENIC AND CHEMOKINETIC
CC       PROPERTIES. HAS A POTENT CHEMOTACTIC ACTIVITY FOR EOSINOPHILS.
CC       BINDING TO A HIGH-AFFINITY RECEPTOR ACTIVATES CALCIUM RELEASE IN
CC       NEUTROPHILS.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN LUNG, SPLEEN, AND PANCREAS.
CC   -!- SIMILARITY: BELONGS TO THE INTERCRINE BETA FAMILY (SMALL CYTOKINE
CC       C-C) (CHEMOKINE CC).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M23447; G533241; -.
DR   EMBL; X12531; G53123; -.
DR   EMBL; X53372; G297531; -.
DR   EMBL; J04491; G201525; -.
DR   EMBL; M73061; G199695; -.
DR   EMBL; AF065939; G3158440; -.
DR   EMBL; AF065940; G3158442; -.
DR   EMBL; AF065941; G3158444; -.
DR   EMBL; AF065942; G3158446; -.
DR   EMBL; AF065943; G3158448; -.
DR   PIR; A27596; A27596.
DR   PIR; A30552; A30552.
DR   PIR; A32393; A32393.
DR   PIR; S04533; S04533.
DR   PIR; S11685; S11685.
DR   MGD; MGI:98260; SCYA3.
DR   PROSITE; PS00472; SMALL_CYTOKINES_CC; 1.
DR   PFAM; PF00048; il8; 1.
DR   HSSP; P13236; 1HUN.
KW   CYTOKINE; CHEMOTAXIS; INFLAMMATORY RESPONSE; SIGNAL.
FT   SIGNAL       1     23
FT   CHAIN       24     92       MACROPHAGE INFLAMMATORY PROTEIN 1-ALPHA.
FT   DISULFID    34     57       BY SIMILARITY.
FT   DISULFID    35     73       BY SIMILARITY.
FT   CONFLICT    22     22       F -> L (IN REF. 3).
FT   CONFLICT    62     62       V -> A (IN REF. 3).
SQ   SEQUENCE   92 AA;  10345 MW;  53979E5E CRC32;
```

```
  Query Match             10.3%;  Score 73;  DB 1;  Length 92;
  Best Local Similarity   30.8%;  Pred. No. 0.06;
  Matches   16;  Conservative   15;  Mismatches   11;  Indels   10;  Gaps    3;

Qy      30 CCLAYHYPIGWAVLRRAWTYRIQ--EVSGSCNLPAAIFYLPKRHRKVCGNPK 79
           ||::|   |       :: : ::  | |: |: |: || | ||:| :|:::|
Db      34 CCFSYSRKI-------PRQFIVDYFETSSLCSQPGVIF-LTKRNRQICADSK 77
```

```
RESULT  15
LTN_MOUSE
ID   LTN_MOUSE      STANDARD;      PRT;   114 AA.
AC   P47993;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   LYMPHOTACTIN PRECURSOR (CYTOKINE SCM-1) (LYMPHOTAXIN).
GN   SCYC1 OR LTN OR LPTN.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=THYMUS;
RX   MEDLINE; 95064019.
RA   KELNER G.S., ZLOTNIK A.;
RT   "Lymphotactin: a cytokine that represents a new class of chemokine.";
RL   SCIENCE 266:1395-1399(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=SPLEEN;
RX   MEDLINE; 95180438.
RA   YOSHIDA T., IMAI T., KAKIZAKI M., NISHIMURA M., YOSHIE O.;
RT   "Molecular cloning of a novel C or gamma type chemokine, SCM-1.";
RL   FEBS LETT. 360:155-159(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SVJ; TISSUE=LIVER;
RA   HAUTAMAA D., MERICA R., CHEN Z.Y., JENKINS M.K.;
RL   SUBMITTED (JUN-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: CHEMOTACTIC ACTIVITY FOR LYMPHOCYTES BUT NOT FOR
CC       MONOCYTES OR NEUTROPHILS.
CC   -!- SIMILARITY: BELONGS TO THE INTERCRINE GAMMA FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U15607; G595909; -.
DR   EMBL; D43769; G927657; -.
DR   EMBL; U28493; G896456; -.
DR   EMBL; U28491; G896456; JOINED.
DR   EMBL; U28492; G896456; JOINED.
DR   MGD; MGI:104593; LPTN.
DR   PFAM; PF00048; il8; 1.
DR   HSSP; P80098; 1NCV.
KW   CYTOKINE; CHEMOTAXIS; SIGNAL.
FT   SIGNAL       1      ?       POTENTIAL.
FT   CHAIN        ?    114       LYMPHOCTATIN.
FT   DISULFID    32     69       POTENTIAL.
FT   CONFLICT   110    110       I -> V (IN REF. 1).
SQ   SEQUENCE   114 AA;  12467 MW;  52D16A00 CRC32;
```

```
  Query Match             10.2%;  Score 72.5;  DB 1;  Length 114;
  Best Local Similarity   23.9%;  Pred. No. 0.088;
  Matches   27;  Conservative   31;  Mismatches   44;  Indels   11;  Gaps    4;

Qy       1 MNLWLLACLVAGFLGAWAPAVHAQGVFED-CCLAYHYPIGWAVLRRAWTYRIQEVSGSCN 59
           | | ||: |    | :|     :  |:|:  |:  :    :  :::  || | | :
Db       1 MRLLLLTFLGVCCLTPWVVEGVGTEVLEESSCVNLQ--TQRLPVQKIKTYIIWEGAMR-- 57

Qy      60 LPAAIFYLPKRHRKVCGNPKSREVQRAMKLLDARNKVFAKLRHNTQTFQAGPH 112
              |: :: ||  |:|::| :: |  |:| :|:|    :  :: ::|  :|::
Db      57 ---AVIFVTKRGLKICADPEAKWVKAAIKTVDGRA---STRKNMAETVPTGAQ 103
```

```
Search completed: July 24, 1999, 10:26:45
Job time: 11853 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 24, 1999, 07:14:08 ; Search time 282.39 Seconds
(without alignments)
6929.258 Million cell updates/sec

Title: US-09-002-485-102
Perfect score: 992
Sequence: 1 GACAGCTTGGCCTACAGCCC.....TGATGGAGTCAAAAAAAAAA 992

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	467.2	47.1	523	44	AI313153	AI313153 qp81e03.x
c 2	440.8	44.4	458	42	AI091662	AI091662 oo19h06.x
c 3	357.6	36.0	407	25	N73958	N73958 za74f04.s1
c 4	320	32.3	345	44	AI313430	AI313430 qp80d12.x
5	286	28.8	291	31	AA295814	AA295814 EST101116
6	263.4	26.6	279	31	AA295958	AA295958 EST101182
7	259.8	26.2	274	31	AA295925	AA295925 EST101148
8	248.2	25.0	256	31	AA295945	AA295945 EST101169
c 9	243.4	24.5	245	43	AI185600	AI185600 qe55d01.x
10	178.2	18.0	379	30	Z84185	Z84185 SSZ84185 Po
11	170.4	17.2	444	31	AA277792	AA277792 vb68a11.r
12	135.4	13.6	530	37	AA674898	AA674898 vq57b12.r
13	132	13.3	378	26	W91616	W91616 MTA.F02.091
14	39.2	4.0	445	47	AI491140	AI491140 EST241849
c 15	37.6	3.8	565	47	AI536479	AI536479 vu59a05.x
c 16	36.4	3.7	323	39	AA830441	AA830441 oc51c07.s
c 17	36.2	3.6	263	39	AA845062	AA845062 ak58d07.s
18	35	3.5	509	27	AA031820	AA031820 zk14a11.r
19	35	3.5	426	29	AA193846	AA193846 rs04f10.r
20	35	3.5	446	29	AA193869	AA193869 rs04c11.r
21	34.6	3.5	714	36	AA620145	AA620145 vo63a09.r
22	34.6	3.5	641	36	AA637600	AA637600 vr28h08.r
23	34.6	3.5	776	38	AA795931	AA795931 vu05g06.r
24	34.6	3.5	538	41	AI019845	AI019845 ua91h08.r
c 25	34.6	3.5	718	41	AI052545	AI052545 oz27h07.x
26	34.6	3.5	772	44	AI323202	AI323202 mp63c06.y
27	34.4	3.5	346	25	D80733	D80733 HUM100F10B
c 28	34.4	3.5	712	51	AU070267	AU070267 AU070267
c 29	34.2	3.4	328	33	AA388051	AA388051 vc86e06.r
30	34.2	3.4	433	37	AA690538	AA690538 vt24c05.r
31	34.2	3.4	441	37	AA725516	AA725516 ai18g02.s
c 32	34	3.4	435	21	R16177	R16177 ya48a02.r1
c 33	34	3.4	494	27	AA002153	AA002153 zh81b01.s
c 34	34	3.4	438	28	C19276	C19276 C19276 Rice
35	34	3.4	312	30	AA263160	AA263160 PMY0542 K
c 36	33.8	3.4	448	25	N53038	N53038 yv53d09.s1
c 37	33.8	3.4	471	29	AA128030	AA128030 zl14a05.r
c 38	33.8	3.4	542	46	AI411117	AI411117 EST239411
c 39	33.8	3.4	744	50	AI686615	AI686615 tx08h11.x
40	33.8	3.4	410	51	AI717799	AI717799 tenf0854
41	33.6	3.4	316	36	C73426	C73426 C73426 Rice
42	33.6	3.4	410	39	AA820094	AA820094 LD22269.5
43	33.6	3.4	416	39	AA820097	AA820097 LD22280.5
c 44	33.6	3.4	258	40	AA921397	AA921397 ak56g07.s
c 45	33.6	3.4	660	45	AI326342	AI326342 ml61f04.x

ALIGNMENTS

RESULT 1
AI313153/c
LOCUS AI313153 523 bp mRNA EST 14-DEC-1998
DEFINITION qp81e03.x1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone IMAGE:1929436 3' similar to TR:O15444 O15444 CHEMOKINE. ;, mRNA sequence.
ACCESSION AI313153
NID g4018758

```
VERSION     AI313153.1  GI:4018758
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 523)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 29, 1997 this sequence version replaced gi:1520604.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 288.
FEATURES             Location/Qualifiers
     source          1. .523
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1929436"
                     /clone_lib="Soares_fetal_lung_NbHL19W"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAATTTTTTTTTTTTTTTTT-3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. This library was constructed
                     from the same fetus as the fetal heart library, Soares
                     fetal heart NbHH19W."
BASE COUNT      132 a    119 c    157 g    115 t
ORIGIN
```

```
  Query Match             47.1%;  Score 467.2;  DB 44;  Length 523;
  Best Local Similarity   94.5%;  Pred. No. 1e-118;
  Matches  495;  Conservative    0;  Mismatches   28;  Indels    1;  Gaps    1;

Qy     463 GGAAACTCCAAGTTATCATCATCCAAGTTTAGCAATCCCATCAGCAGCAGCAAGAGGAAT 522
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     523 GGAAACTCCAAGTTATCATCATCCAAGTTTAGCAATCCCATCAGCAGCAGCAAGAGGAAT 464

Qy     523 GTCTCCCTCCTGATATCAGCTAATTCAGGACTGTGAGCCGGCTCATTTCTGGGCTCCATC 582
           |||||||| ||    |   ||||||||||||||||||| |||||||||||||||||| ||||
Db     463 GTCTCCCT-CTATCACACGCTAATTCAGGACTGTGACCCGGCTCATTTCTGGGCTACATC 405

Qy     583 GGCACAGGAGGGGCCGGATCTTTCTCCGATAAAACCGTCGCCCTACAGACCCAGCTGTCC 642
           ||||||||||||||| |||||||||||||||||| | |||| |||||||| ||| ||||||
Db     404 GGCACAGGAGGGGCAGGATCTTTCTCCGATAAATCAGTCGGCCTACAGAGCCACCTGTCC 345

Qy     643 CCACGCCTCTGTCTTTTGGGTCAAGTCTTAATCCCTGCACCTGAGTTGGTCCTCCCTCTG 702
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     344 CCACGCGTCTGTCTTTTGGGTCAAGTCTTAATCCCTGCACCTGAGTTGGTCCTCCCTCTG 285

Qy     703 CACCCCCACCACCTCCTGCCCGTCTGGCAACTGGAAAGAGGGAGTTGGCCTGATTTTAAG 762
           |||||||||||||||||||||||||||||||||||||||||| ||||||| ||||||||||
Db     284 CACCCCCACCACCTCCTGCCCGTCTGGCAACTGGAAAGAGGAAGTTGGCGTGATTTTAAG 225

Qy     763 CCTTTTGCCGCTCCGGGGACCAGCAGCAATCCTGGGCAGCCAGTGGCTCTTGTAGAGAAG 822
           |||||||||||||||||||||||||||||||||||||  ||||||||||||||| |||
Db     224 CCTTTTGCCGCTCCGGGGACCAGCAGCAATCCTGGGCACACAGTGGCTCTTGTAGACAAG 165

Qy     823 ACTTAGGATACCTCTCTCACTTTCTGTTTCTTGCCGTCCACCCCGGGCCATGCCAGTGTG 882
           |||||||||||||||||||||||| ||||||  ||||||||||||||||||||||||||||
Db     164 ACTTAGGATACCTCTCTCACTTTGTGTTTCAGGCCGTCCACCCCGGGCCATGCCAGTGTG 105

Qy     883 TCCCTCTGGGTCCCTCCAAAACTCTGGTCAGTTCAAGGATGCCCCTCCCAGGCTATGCTT 942
           |||||||||||||||||||||||||||||||||||||||| ||||||||| |||| |||||||
Db     104 TCCCTCTGGGTCCCTCCAAAACTCTGGTCAGTTCAAGCATGCCCCTCACAGGATATGCTT 45

Qy     943 TTCTATAACTTTTAAATAAACCTTGGGGGTTGATGGAGTCAAAA 986
           ||||||| |||||||||||||||||||| ||||||||||||||
Db      44 TTCTATATCTTTTAAATAAACCTTGGGGGGTGATGGAGTCAAAA 1
```

```
RESULT   2
AI091662/c
LOCUS       AI091662      458 bp    mRNA            EST       13-FEB-1999
DEFINITION  oo19h06.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
            IMAGE:1566683 3', mRNA sequence.
ACCESSION   AI091662
NID         g3430721
VERSION     AI091662.1  GI:3430721
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 458)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 17, 1998 this sequence version replaced gi:1901070.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1064   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 420.
FEATURES             Location/Qualifiers
     source          1. .458
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1566683"
                     /clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from five normalized
                     libraries were mixed, and ss circles were made in vitro.
                     Following HAP purification, this DNA was used as tracer in
                     a subtractive hybridization reaction. The driver was
                     PCR-amplified cDNAs from pools of 5,000 clones made from
                     the same 5 libraries. The pools consisted of the following
                     libraries and cloneIDs: Soares NbHSF pool 1:
                     309384-310919, 323208-325895 Soares Nb2HP pool 1:
                     145032-147335, 147720-148103, 148872-149255, 15002 -
                     150407, 151176-152327 Soares Nb2HF8-9W pool 1:
                     758280-760583, 772104-774407 Soares NbHPA pool 1:
                     304776-306311, 320136-322823, 326280-326663 Soares NbHOT
                     pool 1:  723720-726407, 739080-740999 Subtraction by Bento
                     Soares and M. Fatima Bonaldo."
BASE COUNT      113 a    106 c    137 g    102 t
ORIGIN
```

```
  Query Match             44.4%;  Score 440.8;  DB 42;  Length 458;
  Best Local Similarity   99.5%;  Pred. No. 1.8e-111;
  Matches  442;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy     549 AGGACTGTGAGCCGGCTCATTTCTGGGCTCCATCGGCACAGGAGGGGCCGGATCTTTCTC 608
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    458 AGGACTGTGAGCCGGCTCATTTCTGGGCTCCATCGGCACAGGAGGGGCCGGATCTTTCTC 399

Qy    609 CGATAAAACCGTCGCCCTACAGACCCAGCTGTCCCCACGCCTCTGTCTTTTGGGTCAAGT 668
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    398 CGATAAAACCGTCGCCCTACAGACCCAGCTGTCCCCACGCCTCTGTCTTTTGGGTCAAGT 339

Qy    669 CTTAATCCCTGCACCTGAGTTGGTCCTCCCTCTGCACCCCCACCACCTCCTGCCCGTCTG 728
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    338 CTTAATCCCTGCACCTGAGTTGGTCCTCCCTCTGCACCCCCACCACCTCCTGCCCGTCTG 279

Qy    729 GCAACTGGAAAGAGGGAGTTGGCCTGATTTTAAGCCTTTTGCCGCTCCGGGGACCAGCAG 788
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    278 GCAACTGGAAAGAGGGAGTTGGCCTGATTTTAAGCCTTTTGCCGCTCCGGGGACCAGCAG 219

Qy    789 CAATCCTGGGCAGCCAGTGGCTCTTGTAGAGAAGACTTAGGATACCTCTCTCACTTTCTG 848
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    218 CAATCCTGGGCAGCCAGTGGCTCTTGTAGAGAAGACTTAGGATACCTCTCTCACTTTCTG 159

Qy    849 TTTCTTGCCGTCCACCCCGGGCCATGCCAGTGTGTCCCTCTGGGTCCCTCCAAAACTCTG 908
          ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    158 TTTCTTGCCGTCCACCCCGGGCCATGCCAGTGTGTCCCTTTGGGTCCCTCCAAAACTCTG 99

Qy    909 GTCAGTTCAAGGATGCCCCTCCCAGGCTATGCTTTTCTATAACTTTTAAATAAACCTTGG 968
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     98 GTCAGTTCAAGGATGCCCCTCCCAGGCTATGCTTTTCTATAACTTTTAAATAAACCTTGG 39

Qy    969 GGGTTGATGGAGTCAAAAAAAAAA 992
          ||| ||||||||||||||||||||
Db     38 GGGGTGATGGAGTCAAAAAAAAAA 15
```

RESULT 3
N73958/c
LOCUS N73958 407 bp mRNA EST 19-MAR-1996
DEFINITION za74f04.s1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone IMAGE:298303 3', mRNA sequence.
ACCESSION N73958
NID g1231243
VERSION N73958.1 GI:1231243
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 407)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On Nov 22, 1995 this sequence version replaced gi:1071087.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: m13 -40 forward
High quality sequence stop: 320.

FEATURES Location/Qualifiers
source 1. .407
/organism="Homo sapiens"
/db_xref="GDB:1243227"
/db_xref="taxon:9606"
/clone="IMAGE:298303"
/clone_lib="Soares_fetal_lung_NbHL19W"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: lung; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAATTTTTTTTTTTTTTTT-3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal heart library, Soares fetal heart NbHH19W."
BASE COUNT 105 a 95 c 130 g 75 t 2 others
ORIGIN

Query Match 36.0%; Score 357.6; DB 25; Length 407;
Best Local Similarity 98.0%; Pred. No. 1.5e-88;
Matches 393; Conservative 0; Mismatches 5; Indels 3; Gaps 3;

```
Qy    583 GGCACAGGAGGGGCCGGATCTTTCTCCGATAAAACCGTCGCCCTACAGACCCAGCTGTCC 642
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    399 GCCACAGGAGGGGCCGGATCTTTCTCCGATAAAACCGTCGCCCTACAGACCCAGCTGTCC 340

Qy    643 CCACGCCTCTGTCTTTTGGGTCAAGTCTTAATCCCTGCACCTGAGTTGGTCCTCCCTCTG 702
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    339 CCACGCCTCTGTCTTTTGGGTCAAGTCTTAATCCCTGCACCTGAGTTGGTCCTCCCTCTG 280

Qy    703 CACCCCCACCACCTCCTGCCCGTCTGGCAACTGGAAAGAGGGAGTTGGCCTGATTTTAAG 762
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    279 CACCCCCACCACCTCCTGCCCGTCTGGCAACTGGAAAGAGGGAGTTGGCCTGATTTTAAG 220

Qy    763 CCTTTTGCCGCTCCGGGGACCAGCAGCAATCCTGGGCAGCCAGTGGCTCTTGTAGAGAAG 822
          ||||||||||||||||||||||||||||||||||   ||||||||||||||||||||||
Db    219 CCTTTTGCCGCTCCGGGGACCAGCAGCAATCCTGGGCGANCAGTGGCTCTTGTAGAGAAG 160

Qy    823 ACTTAGGATACCTCTCTCACTTTCTGTTTCTTGCCGTCCACCCCGGGCCATGCCAGTGTG 882
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    159 ACTTAGGATACCTCTCTCACTTTCTGTTTCTTGCCGTCCACCCC-GGCCATGCCAGTGTG 101

Qy    883 TCCCTCTGGGTCCCTCCAAAACTCTGGTCAGTTCAAGGATGCCCCTCCC-AGGCTATGCT 941
          | ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db    100 T-CCTCTGGGTCCCTCCAAAACTCTGGTCAGTTCAAGGATGCCCCTCCCNAGGCTATGCT 42

Qy    942 TTTCTATAACTTTTAAATAAACCTTGGGGGTTGATGGAGTC 982
          |||||||||||||||||||||||||||| ||||||||||
Db     41 TTTCTATAACTTTTAAATAAACCTTGGGGGGTGATGGAGTC 1
```

RESULT 4
AI313430/c
LOCUS AI313430 345 bp mRNA EST 03-FEB-1999
DEFINITION qp80d12.x1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone IMAGE:1929335 3', mRNA sequence.
ACCESSION AI313430
NID g4019035
VERSION AI313430.1 GI:4019035
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 345)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

```
            Insert Length: 1053   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 341.
FEATURES              Location/Qualifiers
     source           1. .345
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1929335"
                      /clone_lib="Soares_fetal_lung_NbHL19W"
                      /dev_stage="19 weeks"
                      /lab_host="DH10B (ampicillin resistant)"
                      /note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
                      modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer
                      [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAATTTTTTTTTTTTTTTTT-3'],
                      double-stranded cDNA was size selected, ligated to Eco RI
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pT7T3 vector
                      (Pharmacia). Library went through one round of
                      normalization to a Cot = 5. Library constructed by Bento
                      Soares and M.Fatima Bonaldo. This library was constructed
                      from the same fetus as the fetal heart library, Soares
                      fetal heart NbHH19W."
BASE COUNT       97 a     79 c    104 g     63 t      2 others
ORIGIN


  Query Match             32.3%;  Score 320;  DB 44;  Length 345;
  Best Local Similarity   97.9%;  Pred. No. 3.3e-78;
  Matches  323;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy     654 TCTTTTGGGTCAAGTCTTAATCCCTGCACCTGAGTTGGTCCTCCCTCTGCACCCCCACCA 713
           ||||||||||||||||||||||||||||||   | ||||||||||||||||||  |||
Db     345 TCTTTTGGGTCAAGTCTTAATCCCTGCACCTAGTTNGGTCCTCCCTCTGCACCCCANCCA 286

Qy     714 CCTCCTGCCCGTCTGGCAACTGGAAAGAGGGAGTTGGCCTGATTTTAAGCCTTTTGCCGC 773
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     285 CCTCCTGCCCGTCTGGCAACTGGAAAGAGGGAGTTGGCCTGATTTTAAGCCTTTTGCCGC 226

Qy     774 ICCGGGGACCAGCAGCAATCCTGGGCAGCCAGTGGCTCTTGTAGAGAAGACTTAGGATAC 833
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     225 TCCGGGGACCAGCAGCAATCCTGGGCAGCCAGTGGCTCTTGTAGAGAAGACTTAGGATAC 166

Qy     834 CTCTCTCACTTTCTGTTTCTTGCCGTCCACCCCGGGCCATGCCAGTGTGTCCCTCTGGGT 893
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     165 CTCTCTCACTTTCTGTTTCTTGCCGTCCACCCCGGGCCATGCCAGTGTGTCCCTCTGGGT 106

Qy     894 CCCTCCAAAACTCTGGTCAGTTCAAGGATGCCCCTCCCAGGCTATGCTTTTCTATAACTT 953
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     105 CCCTCCAAAACTCTGGTCAGTTCAAGGATGCCCCTCCCAGGCTATGCTTTTCTATAACTT 46

Qy     954 TTAAATAAACCTTGGGGGTTGATGGAGTCA 983
           |||||||||||||||| ||||||||||
Db      45 TTAAATAAACCTTGGGGGGTGATGGAGTCA 16


RESULT   5
AA295814
LOCUS       AA295814      291 bp    mRNA            EST       18-APR-1997
DEFINITION  EST101116 Thymus III Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA295814
NID         g1948238
VERSION     AA295814.1  GI:1948238
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     On Apr 14, 1993 this sequence version replaced gi:692368.
            Other_ESTs: THC148619
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .291
                      /organism="Homo sapiens"
                      /db_xref="ATCC (inhost):191341"
                      /db_xref="taxon:9606"
                      /clone_lib="Thymus III"
                      /dev_stage="adult"
                      /note="Organ: thymus; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI"
BASE COUNT       48 a     98 c     64 g     76 t      5 others
ORIGIN


  Query Match             28.8%;  Score 286;  DB 31;  Length 291;
  Best Local Similarity   98.3%;  Pred. No. 7.4e-69;
  Matches  286;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy     601 TCTTTCTCCGATAAAACCGTCGCCCTACAGACCCAGCTGTCCCCACGCCTCTGTCTTTTG 660
           |||||||||||||||||||||||||||||||||||||||||||||||||||| | |||||
Db       1 TCTTTCTCCGATAAAACCGTCGCCCTACAGACCCAGCTGTCCCCACGCCTCTNTNTTTTG 60

Qy     661 GGTCAAGTCTTAATCCCTGCACCTGAGTTGGTCCTCCCTCTGCACCCCCACCACCTCCTG 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GGTCAAGTCTTAATCCCTGCACCTGAGTTGGTCCTCCCTCTGCACCCCCACCACCTCCTG 120

Qy     721 CCCGTCTGGCAACTGGAAAGAGGGAGTTGGCCTGATTTTAAGCCTTTTGCCGCTCCGGGG 780
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CCCGTNTGGCAACTGGAAAGAGGGAGTTGGCCTGATTTTAAGCCTTTTGCCGCTCCGGGG 180

Qy     781 ACCAGCAGCAATCCTGGGCAGCCAGTGGCTCTTGTAGAGAAGACTTAGGATACCTCTCTC 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ACCAGCAGCAATCCTGGGCAGCCAGTGGCTCTTGTAGAGAAGACTTAGGATACCTCTCTC 240

Qy     841 ACTTTCTGTTTCTTGCCGTCCACCCCGGGCCATGCCAGTGTGTCCCTCTGG 891
           |||||||||||||||||||||||||||| ||||||||| ||||||||||
Db     241 ACTTTCTGTTTCTTGCCGTCCACCCCGGGNCATGCCAGTNTGTCCCTCTGG 291


RESULT   6
AA295958
LOCUS       AA295958      279 bp    mRNA            EST       18-APR-1997
DEFINITION  EST101182 Thymus III Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA295958
```

```
NID         g1948303
VERSION     AA295958.1  GI:1948303
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 279)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     On Dec 30, 1996 this sequence version replaced gi:1529533.
            Other_ESTs: THC148619
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .279
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):191405"
                     /db_xref="taxon:9606"
                     /clone_lib="Thymus III"
                     /dev_stage="adult"
                     /note="Organ: thymus; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"
BASE COUNT       48 a     85 c     62 g     78 t      6 others
ORIGIN
```

```
  Query Match             26.6%;  Score 263.4;  DB 31;  Length 279;
  Best Local Similarity   95.7%;  Pred. No. 1.2e-62;
  Matches  267;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy    645 ACGCCTCTGTCTTTTGGGTCAAGTCTTAATCCCTGCACCTGAGTTGGTCCTCCCTCTGCA 704
          |||||||| | |||||| ||||| |||||||||||||| | ||||||||||||||||
Db      1 ACGCCTCTTTTTTTTGGNTCAAGTNTTAATCCCTGCACCTAANTTGGTCCTCCCTCTGCA 60

Qy    705 CCCCCACCACCTCCTGCCCGTCTGGCAACTGGAAAGAGGGAGTTGGCCTGATTTTAAGCC 764
          ||||||||||||||||||| |||||||||||||||||||||||||||||||| |||||
Db     61 CCCCCACCACCTCCTGCCCGTTTGGCAACTGGAAAGAGGGAGTTGGCCTGATTTNAAGCC 120

Qy    765 TTTTGCCGCTCCGGGGGACCAGCAGCAATCCTGGGCAGCCAGTGGCTCTTGTAGAGAAGAC 824
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TTTTGCCGCTCCGGGGACCAGCAGCAATCCTGGGCAGCCAGTGGCTCTTGTAGAGAAGAC 180

Qy    825 TTAGGATACCTCTCTCACTTTCTGTTTCTTGCCGTCCACCCCGGGCCATGCCAGTGTGTC 884
          |||||||||||||| | ||||||||||||||||||||||||||||||||||||| ||||
```

```
Db    181 TTAGGATACCTCTNTNACTTTCTGTTTCTTGCCGTCCACCCCGGGCCATGCCAGTTTGTC 240

Qy    885 CCTCTGGGTCCCTCCAAAACTCTGGTCAGTTCAAGGATG 923
          ||||||||||||||||||| ||||||||||||||||
Db    241 CCTCTGGGTCCCTCCAAAACTTTGGTCAGTTCAAGGATG 279
```

```
RESULT   7
AA295925
LOCUS       AA295925      274 bp    mRNA            EST       18-APR-1997
DEFINITION  EST101148 Thymus III Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA295925
NID         g1948269
VERSION     AA295925.1  GI:1948269
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 274)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     On Dec 30, 1996 this sequence version replaced gi:1529576.
            Other_ESTs: THC148619
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .274
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):191371"
                     /db_xref="taxon:9606"
                     /clone_lib="Thymus III"
                     /dev_stage="adult"
                     /note="Organ: thymus; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"
BASE COUNT       60 a     79 c     57 g     71 t      7 others
ORIGIN
```

```
  Query Match             26.2%;  Score 259.8;  DB 31;  Length 274;
  Best Local Similarity   96.7%;  Pred. No. 1.2e-61;
  Matches  261;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy    435 CCCTCATGCTGTAAAGAAGTTGAGTTCTGGAAACTCCAAGTTATCATCATCCAAGTTTAG 494
          |||||||||||||||||||| |||||||||||||||||||||||||| ||||||||||
```

```
Db      1 CCCTCATGCTGTAAAGAAGTTNAGTTCTGGAAACTCCAAGTTATCATCGTCCAAGTTTAG 60

Qy    495 CAATCCCATCAGCAGCAGCAAGAGGAATGTCTCCCTCCTGATATCAGCTAATTCAGGACT 554
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db     61 CAATCCCATCAGCAGCAGCAAGAGGAATGTNTCCCTCCTGATATCAGCTAATTCAGGACT 120

Qy    555 GTGAGCCGGCTCATTTCTGGGCTCCATCGGCACAGGAGGGGCCGGATCTTTCTCCGATAA 614
          || |||||||||||| |||||||||||||||||||||||||||||||| ||||||||
Db    121 GTAAGCCGGCTCATTTNTGGGCTCCATCGGCACAGGAGGGGCCGGATCTTTNTCCGATAA 180

Qy    615 AACCGTCGCCCTACAGACCCAGCTGTCCCCACGCCTCTGTCTTTTGGGTCAAGTCTTAAT 674
          |||||||||||||||||||||||||||||| |||| | ||||||||||||||||||||
Db    181 AACCGTCGCCCTACAGACCCAGCTGTCCCCACGNCTCTNTNTTTTGGGTCAAGTCTTAAT 240

Qy    675 CCCTGCACCTGAGTTGGTCCTCCCTCTGCA 704
          ||||||||||||||||||||||||||||||
Db    241 CCCTGCACCTGAGTTGGTCCTCCCTCTGCA 270


RESULT   8
AA295945
LOCUS       AA295945      256 bp    mRNA            EST       18-APR-1997
DEFINITION  EST101169 Thymus III Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA295945
NID         g1948290
VERSION     AA295945.1  GI:1948290
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 256)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     On Dec 30, 1996 this sequence version replaced gi:1529604.
            Other_ESTs: THC148619
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .256
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):191392"
                     /db_xref="taxon:9606"
                     /map="19"
                     /clone_lib="Thymus III"
                     /dev_stage="adult"
                     /note="Organ: thymus; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"
BASE COUNT       70 a     66 c     61 g     56 t      3 others
ORIGIN


  Query Match             25.0%;  Score 248.2;  DB 31;  Length 256;
  Best Local Similarity   97.7%;  Pred. No. 1.9e-58;
  Matches  250;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy    329 GTGGGAACCCCAAAAGCAGGGAGGTGCAGAGAGCCATGAAGCTCCTGGATGCTCGAAATA 388
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db      1 GTGGGAACCCCAAAAGCAGGGAGGTGCAGAGAGCCATGAAGCTCCTGGNTGCTCGAAATA 60

Qy    389 AGGTTTTTGCAAAGCTCCGCCACAACACGCAGACCTTCCAAGCAGGCCCTCATGCTGTAA 448
          |||||||| ||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     61 AGGTTTTTNCAAAGCTCCACCACAACACGCAGACCTTCCAAGCAGGCCCTCATGCTGTAA 120

Qy    449 AGAAGTTGAGTTCTGGAAACTCCAAGTTATCATCATCCAAGTTTAGCAATCCCATCAGCA 508
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    121 AGAAGTTGAGTTCTGGAAACTCCAAGTTATCATCGTCCAAGTTTAGCAATCCCATCAGCA 180

Qy    509 GCAGCAAGAGGAATGTCTCCCTCCTGATATCAGCTAATTCAGGACTGTGAGCCGGCTCAT 568
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    181 GCAGCAAGAGGAATGTTTCCCTCCTGATATCAGCTAATTCAGGACTGTGAGCCGGCTCAT 240

Qy    569 TTCTGGGCTCCATCGG 584
          || ||||||||||||||
Db    241 TTNTGGGCTCCATCGG 256


RESULT   9
AI185600/c
LOCUS       AI185600      245 bp    mRNA            EST       29-OCT-1998
DEFINITION  qe55d01.x1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
            IMAGE:1742881 3', mRNA sequence.
ACCESSION   AI185600
NID         g3736238
VERSION     AI185600.1  GI:3736238
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 245)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2151085.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1035   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 235.
FEATURES             Location/Qualifiers
     source          1. .245
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="960H11; 6; 6p21.31-6p22.1; 21q"
                     /clone="IMAGE:1742881"
                     /clone_lib="Soares_fetal_lung_NbHL19W"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAATTTTTTTTTTTTTTTT-3'],
```

double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal heart library, Soares fetal heart NbHH19W."

```
BASE COUNT       68 a     61 c     68 g     48 t
ORIGIN
```

```
  Query Match             24.5%;  Score 243.4;  DB 43;  Length 245;
  Best Local Similarity   99.6%;  Pred. No. 3.9e-57;
  Matches  244;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    741 AGGGAGTTGGCCTGATTTTAAGCCTTTTGCCGCTCCGGGGACCAGCAGCAATCCTGGGCA 800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    245 AGGGAGTTGGCCTGATTTTAAGCCTTTTGCCGCTCCGGGGACCAGCAGCAATCCTGGGCA 186

Qy    801 GCCAGTGGCTCTTGTAGAGAAGACTTAGGATACCTCTCTCACTTTCTGTTTCTTGCCGTC 860
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    185 GCCAGTGGCTCTTGTAGAGAAGACTTAGGATACCTCTCTCACTTTCTGTTTCTTGCCGTC 126

Qy    861 CACCCCGGGCCATGCCAGTGTGTCCCTCTGGGTCCCTCCAAAACTCTGGTCAGTTCAAGG 920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    125 CACCCCGGGCCATGCCAGTGTGTCCCTCTGGGTCCCTCCAAAACTCTGGTCAGTTCAAGG 66

Qy    921 ATGCCCCTCCCAGGCTATGCTTTTCTATAACTTTTAAATAAACCTTGGGGGGTTGATGGAG 980
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db     65 ATGCCCCTCCCAGGCTATGCTTTTCTATAACTTTTAAATAAACCTTGGGGGGTGATGGAG 6

Qy    981 TCAAA 985
          |||||
Db      5 TCAAA 1
```

```
RESULT  10
Z84185
LOCUS       Z84185        379 bp    mRNA            EST       30-JAN-1997
DEFINITION  SSZ84185 Porcine small intestine cDNA library Sus scrofa cDNA clone
            c15g05 5', mRNA sequence.
ACCESSION   Z84185
NID         g1806525
VERSION     Z84185.1  GI:1806525
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 379)
  AUTHORS   Wintero,A.K., Fredholm,M. and Davies,W.
  TITLE     Evaluation and characterization of a porcine small intestine cDNA
            library: analysis of 839 clones
  JOURNAL   Mamm. Genome 7 (7), 509-517 (1996)
  MEDLINE   96327607
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1404694.

            Contact: A.K. Winteroe
            Department of Animal Science and Animal Health, Division of Animal
            Genetics, The Royal Veterinary and Agricultural University
            Bulowsvej 13, 1870 Frederiksberg C, Denmark.
FEATURES             Location/Qualifiers
     source          1. .379
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone="c15g05"
                     /clone_lib="Porcine small intestine cDNA library"
                     /note="directionally cloned cDNA in XL1-blue MRF'"
BASE COUNT       73 a    105 c    117 g     79 t      5 others
ORIGIN
```

```
  Query Match             18.0%;  Score 178.2;  DB 30;  Length 379;
  Best Local Similarity   79.8%;  Pred. No. 4.4e-39;
  Matches  221;  Conservative    0;  Mismatches   55;  Indels    1;  Gaps    1;

Qy     63 CGTTATTCGTCCAGGTGCCCAGGGAGGAGGACCCGCCTGCAGCATGAACCTGTGGCTCCT 122
          | || ||  |||  | | |   |||||||||| |||||||  ||||  | |||||||||
Db     40 CTTTGTTAATCCTCGAGGCGCCGGAGGAGGACCAGCCTGCACTATGAGGCCGTGGCTCCT 99

Qy    123 GGCCTGCCTGGTGGCCGGCTTCCTGGGAGCCTGGGCCCCCGCTGTCCACGCCCAAGGTGT 182
          ||||||||||||||| ||||| |||| |||||||| || ||| |||| || |||||||
Db    100 GGCCTGCCTGGTGGCCTGCTTCGTGGGTGCCTGGGCTCCTGCTATCCATGCTCAAGGTGC 159

Qy    183 CTTTGAGGACTGCTGCCTGGCCTACCACTACCCCATTGGGTGGGCTGTGCTCCGGCGCGC 242
          |||||||||||||||||||||||||||  || ||||  ||||    ||||||||||||
Db    160 CTTTGAGGACTGCTGCCTGGCCTACCACAGCCACATTAAGTGGCGCCTGCTCCGGCGCGC 219

Qy    243 CTGGACTTACCGGATCCAGGAGGTGAGCGGGAGCTGCAATCTGCCT-GCTGCGATATTCT 301
          |   | ||||| |  |||||| |||||||||| |||||| |||||| |||| ||||||||
Db    220 CCACAGTTACCAGCGCCAGGATGTGAGCGGGANCTGCAACCTGCCTGGCTGTGATATTCT 279

Qy    302 ACCTCCCCAAGAGACACAGGAAGGTGTGTGGGAACCC 338
          | ||||| ||| | ||| || |||||| |||| ||
Db    280 TCTTCCCCCAGAAAGACAAGATGGTGTGCGGGAANCC 316
```

```
RESULT  11
AA277792
LOCUS       AA277792      444 bp    mRNA            EST       01-APR-1997
DEFINITION  vb68a11.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA
            clone IMAGE:762140 5', mRNA sequence.
ACCESSION   AA277792
NID         g1917791
VERSION     AA277792.1  GI:1917791
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 444)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Dec 30, 1996 this sequence version replaced gi:1282605.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:463060
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 403.
FEATURES             Location/Qualifiers
     source          1. .444
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /map="17"
                     /clone="IMAGE:762140"
                     /clone_lib="Barstead mouse pooled organs MPLRB4"
                     /sex="mixed"
                     /tissue_type="pooled organs"
                     /dev_stage="7 day"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
```

a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [GTTGGATTCGGTACC], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed by Bob Barstead."

BASE COUNT 119 a 100 c 124 g 101 t
ORIGIN

Query Match 17.2%; Score 170.4; DB 31; Length 444;
Best Local Similarity 67.1%; Pred. No. 6.7e-37;
Matches 259; Conservative 0; Mismatches 121; Indels 6; Gaps 1;

```
Qy       7 TTGGCCTACAGCCCGGCGGGCATCAGCTCCCTTGACCCAGTGGATATCGGTGGCCCCGTT 66
Db       3 TTCGGATCCAACCAAGCAGGCACCAGCTCTCAGGACCAGAAAGGCATTGGTGGCCCCCTT 62

Qy      67 ATTCGTCCAGGTGCCCAGGGAGGAGGACCCGCCTGCAGCATGAACCTGTGGCTCCTGGCC 126
Db      63 AAACCTTCAGGTATCTGGAGAGGAGATCTAACCTTCACTATGAAACTGTGGCTTTTTGCC 122

Qy     127 TGCCTGGTGGCCGGCTTCCTGGGAGCCTGGGCCCCCGCTGTCCACGCCCAAGGTGTCTTT 186
Db     123 TGCCTGGTTGCCTGTTTTGTTGGGGCCTGGATGCCGGTTGTCCATGCCCAAGGTGCCTTT 182

Qy     187 GAGGACTGCTGCCTGGCCTACCACTACCCCATTGGGTGGGCTGTGCTCCGGCGCGCCTGG 246
Db     183 GAAGACTGCTGCCTGGGTTACCAGCACAGGATCAAATGGAATGTTCTCCGGCATGCTAGG 242

Qy     247 ACTTACCGGATCCAGGAGGTGAGCGGGAGCTGCAATCTGCCTGCTGCGATATTCTACCTC 306
Db     243 AATTATCACCAGCAGGAAGTGAGTGGAAGCTGCAACCTACGTGCTGTGAGATTCTACTTC 302

Qy     307 CCCAAGAGACACAGGAAGGTGTGTGGGAACCCCAAAAGCAGGGAGGTGCAGAGAGCCATG 366
Db     303 CGCCAGAAA------GTAGTGTGTGGGAATCCAGAGGACATGAATGTGAAGAGGGCGATG 356

Qy     367 AAGCTCCTGGATGCTCGAAATAAGGT 392
Db     357 AGAATCTTGACAGCTAGGAAAAGGCT 382
```

RESULT 12
AA674898
LOCUS AA674898 530 bp mRNA EST 28-NOV-1997
DEFINITION vq57b12.r1 Barstead mouse proximal colon MPLRB6 Mus musculus cDNA clone IMAGE:1106399 5', mRNA sequence.
ACCESSION AA674898
NID g2652135
VERSION AA674898.1 GI:2652135
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 530)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1407053.

Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:604567
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 519.

FEATURES Location/Qualifiers
source 1. .530
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:1106399"
/clone_lib="Barstead mouse proximal colon MPLRB6"
/dev_stage="7 day juvenile"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [AATTCGGATCCTTG], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed by Bob Barstead. "

BASE COUNT 143 a 117 c 150 g 120 t
ORIGIN

Query Match 13.6%; Score 135.4; DB 37; Length 530;
Best Local Similarity 68.8%; Pred. No. 3.2e-27;
Matches 203; Conservative 0; Mismatches 86; Indels 6; Gaps 1;

```
Qy      98 CCTGCAGCATGAACCTGTGGCTCCTGGCCTGCCTGGTGGCCGGCTTCCTGGGAGCCTGGG 157
Db      18 CCTTCACTATGAAACTGTGGCTTTTTGCCTGCCTGGTTGCCTGTTTTGTTGGGGCCTGGA 77

Qy     158 CCCCCGCTGTCCACGCCCAAGGTGTCTTTGAGGACTGCTGCCTGGCCTACCACTACCCCA 217
Db      78 TGCCGGTTGTCCATGCCCAAGGTGCCTTTGAAGACTGCTGCCTGGGTTACCAGCACAGGA 137

Qy     218 TTGGGTGGGCTGTGCTCCGGCGCGCCTGGACTTACCGGATCCAGGAGGTGAGCGGGAGCT 277
Db     138 TCAAATGGAATGTTCTCCGGCATGCTAGGAATTATCACCAGCAGGAAGTGAGTGGAAGCT 197

Qy     278 GCAATCTGCCTGCTGCGATATTCTACCTCCCCAAGAGACACAGGAAGGTGTGTGGGAACC 337
Db     198 GCAACCTACGTGCTGTGAGATTCTACTTCCGCCAGAAA------GTAGTGTGTGGGAATC 251

Qy     338 CCAAAAGCAGGGAGGTGCAGAGAGCCATGAAGCTCCTGGATGCTCGAAATAAGGT 392
Db     252 CAGAGGACATGAATGTGAAGAGGGCGATGAGAATCTTGACAGCTAGGAAAAGGCT 306
```

RESULT 13
W91616
LOCUS W91616 378 bp mRNA EST 09-JUL-1996
DEFINITION MTA.F02.091.A MTA adult mouse thymus library Mus musculus cDNA clone MTA.F02.091 5' end, mRNA sequence.
ACCESSION W91616
NID g1407877
VERSION W91616.1 GI:1407877
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 378)
AUTHORS Nguyen,C., Rocha,D., Granjeaud,S., Bernard,K., Naquet,P. and Jordan,B.R.
TITLE Gene expression in different cell types of the mouse thymus
JOURNAL Unpublished (1996)
COMMENT On Apr 14, 1993 this sequence version replaced gi:785677.

Contact: Jordan BR

```
            Genome Structure and Immune Functions
            Centre d'Immunologie INSERM/CNRS
            Case 906, 13288 MARSEILLE Cedex 9, FRANCE
            Tel: 330491269496
            Fax: 330491269430
            Email: jordan@ciml.univ-mrs.fr
            This sequence was determined at HGMP Resource Centre, Hinxton,
            United Kingdom (M Botcherby)
            Seq primer: T7.
FEATURES              Location/Qualifiers
     source           1. .378
                      /organism="Mus musculus"
                      /strain="C57Bl/6"
                      /db_xref="taxon:10090"
                      /clone="MTA.F02.091"
                      /clone_lib="MTA adult mouse thymus library"
                      /lab_host="MC1061 p3"
                      /note="Vector: pcDNA1; Site_1: Not1; Site_2: EcoR1; The
                      cDNA library was constructed from poly(A)+ RNA of an adult
                      mouse  thymus by oligo-dT primed reverse transcription.
                      cDNA was selected on  gel for size above 800 nucleotides
                      after second strand synthesis, then  directionally cloned
                      after adaptor ligation and restriction enzyme  cleavage
                      into the pcDNA1 vector (Not 1 on polyA side, EcoR 1 on the
                      5'side)."
BASE COUNT       93 a     82 c    108 g     93 t      2 others
ORIGIN


  Query Match             13.3%;  Score 132;  DB 26;  Length 378;
  Best Local Similarity   67.5%;  Pred. No. 2.4e-26;
  Matches  216;  Conservative    0;  Mismatches   97;  Indels    7;  Gaps    2;

Qy      48 GGATATCGGTGGCCCCGTTATTCGTCCAGGTGCCCAGGGAGGAGGACCCGCCTGCAGCAT 107
           ||   | |  | |||| |||  | | |||||  |  | |||||| | |   || ||  ||
Db      12 GGCATTGGTGGCCCCCTTTAAACCTTCAGGTATCTGGAGAGGAG-ATCTANCTTCACTAT 70

Qy     108 GAACCTGTGGCTCCTGGCCTGCCTGGTGGCCGGCTTCCTGGGAGCCTGGGCCCCCGCTGT 167
           ||| ||||||||  | ||||||||||| ||    ||  | || ||||||   || | |||
Db      71 GAAACTGTGGCTTTTTGCCTGCCTGGTTGCNTCTTTTGTTGGGGCCTGGATGCCGGTTGT 130

Qy     168 CCACGCCCAAGGTGTCTTTGAGGACTGCTGCCTGGCCTACCACTACCCCATTGGGTGGGC 227
           ||| ||||||||| |||||| |||||||||||||  |||||  ||   ||    |||
Db     131 CCATGCCCAAGGTGCCTTTGAAGACTGCTGCCTGGGTTACCAGCACAGGATCAAATGGAA 190

Qy     228 TGTGCTCCGGCGCGCCTGGACTTACCGGATCCAGGAGGTGAGCGGGAGCTGCAATCTGCC 287
           ||| |||||||  ||  ||| ||| |     ||||| ||||| || ||||||||| || |
Db     191 TGTTCTCCGGCATGCTAGGAATTATCACCAGCAGGAAGTGAGTGGAAGCTGCAACCTACG 250

Qy     288 TGCTGCGATATTCTACCTCCCCAAGAGACACAGGAAGGTGTGTGGGAACCCCAAAAGCAG 347
           |||||| || |||||||  ||| | ||| |          |||||||||| ||  |   ||
Db     251 TGCTGTGAGATTCTACTTCCGCCAGAAA------GTAGTGTGTGGGAATCCAGAGGACAT 304

Qy     348 GGAGGTGCAGAGAGCCATGA 367
           | | ||| |||| || ||||
Db     305 GAATGTGAAGAGGGCGATGA 324


RESULT  14
AI491140
LOCUS       AI491140      445 bp    mRNA            EST       12-APR-1999
DEFINITION  EST241849 tomato shoot, Cornell Lycopersicon esculentum cDNA clone
            cLEB3B1 similar to L-ascorbate peroxidase, mRNA sequence.
ACCESSION   AI491140
NID         g4386450
VERSION     AI491140.1  GI:4386450
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Potatoe; Lycopersicon.
REFERENCE   1  (bases 1 to 445)
  AUTHORS   van der Hoeven,R.S., Matern,A.L., Vision,T., Chandra,I.,
            Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W.,
            Fraser,C.M., Venter,J.C., Martin,G.B., Giovannoni,J.J. and
            Tanksley,S.D.
  TITLE     Generation of ESTs from tomato shoot meristem
  JOURNAL   Unpublished (1999)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3188621.
            Other_ESTs: EST241848 EST242112
            Contact: David Frisch
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 4366
            Fax: 864 656 4293
            Email: dfrisch@CLEMSON.EDU.
FEATURES              Location/Qualifiers
     source           1. .445
                      /organism="Lycopersicon esculentum"
                      /cultivar="TA496"
                      /db_xref="taxon:4081"
                      /clone="cLEB3B1"
                      /clone_lib="tomato shoot, Cornell"
                      /tissue_type="shoot meristem"
                      /dev_stage="8 week old plants"
                      /lab_host="XLOLR"
                      /note="Vector: pBK_CMV; Site_1: EcoR1; Site_2: Xho1; cLEB
                      - Tomato Shoot Meristem EST Library. Oligo-dT primed cDNA
                      library  made from tomato vegetative shoots including
                      meristems and small  expanding leaves."
BASE COUNT      169 a    102 c     77 g     97 t
ORIGIN


  Query Match             4.0%;  Score 39.2;  DB 47;  Length 445;
  Best Local Similarity   60.2%;  Pred. No. 1;
  Matches   65;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy     383 GAAATAAGGTTTTTGCAAAGCTCCGCCACAACACGCAGACCTTCCAAGCAGGCCCTCATG 442
           ||||| |    ||  ||||  |  | ||| || | |||  ||||||| |||   || | |
Db     156 GAAATGAACCATTAACAAAAATATGTCACCACTCCCAGCTCTTCCAATCAGCATCTTAAG 215

Qy     443 CTGTAAAGAAGTTGAGTTCTGGAAACTCCAAGTTATCATCATCCAAGT 490
           ||  |   ||    | ||| | || || ||||| | | ||| |  |||
Db     216 CTTCAGCAAATCCCAATTCAGAAAGCTTCAAGTGAGCCTCAGCATAGT 263


RESULT  15
AI536479/c
LOCUS       AI536479      565 bp    mRNA            EST       18-MAR-1999
DEFINITION  vu59a05.x1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone
            IMAGE:1195664 3' similar to TR:O00175 O00175 MPIF-2. ;, mRNA
            sequence.
ACCESSION   AI536479
NID         g4450614
VERSION     AI536479.1  GI:4450614
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 565)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     On May 18, 1998 this sequence version replaced gi:3137961.

            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
```

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:642760
This clone was previously sequenced on the 5' end only, this new data is from the 3' end
Possible reversed clone: similarity on wrong strand
High quality sequence stop: 406.

FEATURES Location/Qualifiers
source 1. .565
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1195664"
/clone_lib="Soares mouse mammary gland NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 142 a 140 c 141 g 140 t 2 others
ORIGIN

Query Match 3.8%; Score 37.6; DB 47; Length 565;
Best Local Similarity 55.3%; Pred. No. 3.1;
Matches 73; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```
Qy    258 CCAGGAGGTGAGCGGGAGCTGCAATCTGCCTGCTGCGATATTCTACCTCCCCAAGAGACA 317
          ||||  ||  | ||| |||  |  ||     || | | |  ||| | || ||||||
Db    432 CCAGTTGGCCAACGGCAGCATCTGTCCCAAGGCAGGGGTCATCTTCATCACCAAGAAGGG 373

Qy    318 CAGGAAGGTGTGTGGGAACCCCAAAAGCAGGGAGGTGCAGAGAGCCATGAAGCTCCTGGA 377
          |   ||| | |||    |||||||      |  ||| |||||   |||  |    |||||
Db    372 CCATAAGATCTGTACTGACCCCAAGCTGCTGTGGGTTCAGAGGCACATACAAAAACTGGA 313

Qy    378 TGCTCGAAATAA 389
          |||    || ||
Db    312 TGCCAAGAAAAA 301
```

Search completed: July 24, 1999, 11:28:17
Job time: 15249 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 24, 1999, 07:27:23 ; Search time 39.4 Seconds
(without alignments)
4736.705 Million cell updates/sec

Title: US-09-002-485-102
Perfect score: 992
Sequence: 1 GACAGCTTGGCCTACAGCCC.....TGATGGAGTCAAAAAAAAAA 992

Scoring table: IDENTITY_NUC

Searched: 240622 seqs, 94065609 residues

Database : N_Geneseq_34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		984.8	99.3	1029	1	V45444	Human chemokine ZS
2		939.6	94.7	989	1	V04481	Human chemokine be
3		914.4	92.2	1012	1	V15415	Human thymus expre
4		311.8	31.4	450	1	V45445	Human chemokine ZS
5		165	16.6	1034	1	V15414	Mouse thymus expre
6		40.2	4.1	285	1	V63290	Codon optimised hu
7		40.2	4.1	4256	1	V63305	Nucleotide sequenc
8		40.2	4.1	360	1	V63289	DNA encoding human
9		38.6	3.9	360	1	T03207	Human monocyte che
10		38.6	3.9	360	1	T47229	Human monocyte che
c 11		35.6	3.6	1068	1	T94505	Polyamide hydrolas
c 12		35.6	3.6	1068	1	T94507	Polyamide hydrolas
c 13		35.6	3.6	7193	1	V50431	Streptomyces clavu
14		35.2	3.5	8169	1	V26609	Actinomadura hibis
15		34.6	3.5	44377	1	T78508	Platenolide syntha
16		34.6	3.5	44377	1	T80414	Platenolide syntha
17		33.8	3.4	1229	1	Q38229	Sequence of rye gr
18		33.8	3.4	1376	1	Q56747	Ryegrass Lol pV al
19		33.8	3.4	1229	1	Q85932	cDNA encoding Lol
20		33.6	3.4	662	1	Q05299	Sequence of pAT 74
21		33.6	3.4	3946	1	T93610	Mycobacterium tube
c 22		33.4	3.4	4425	1	T03090	Protein tyrosine-k
c 23		33.4	3.4	9108	1	T03104	Plasmid pRK5.tk1-1
c 24		33.4	3.4	4195	1	T12068	FLT4 receptor tyro
25		33.4	3.4	7742	1	T84745	Bordetella pertuss
26		33.4	3.4	43280	1	T80413	Tylactone synthase
27		33.2	3.3	247	1	Q25260	Prepro region of y
28		33.2	3.3	814	1	Q25259	Cytokine encoded b
29		33.2	3.3	810	1	Q85371	Chemoattractant pr
c 30		33	3.3	2183	1	T88432	Human 5-lipoxygena
c 31		33	3.3	2177	1	T88433	Human 5-lipoxygena
c 32		33	3.3	2195	1	T88434	Human 5-lipoxygena
c 33		33	3.3	2189	1	T88431	Human 5-lipoxygena
34		32.8	3.3	4767	1	V21648	AAV4 genome. Adeno
35		32.8	3.3	1872	1	V21649	AAV4 Rep gene. Ade
36		32.8	3.3	1611	1	V21655	AAV4 Rep 68 gene.
37		32.8	3.3	1872	1	V21656	AAV4 Rep 78 gene.
38		32.6	3.3	270	1	Q94093	Macrophage inflamm
39		32.6	3.3	270	1	T44363	cDNA encoding huma
40		32.6	3.3	719	1	T78167	Human dendritic ce
c 41		32.4	3.3	2937	1	Q64883	Endoglin gene. Nuc
c 42		32.4	3.3	1410	1	T13750	Partial sequence o
c 43		32.4	3.3	2025	1	T13756	Endoglin cDNA sequ
44		32.4	3.3	1422	1	T42466	Fibroblast growth
45		32.4	3.3	1422	1	T86314	Human fibroblast g

ALIGNMENTS

RESULT 1
V45444
ID V45444 standard; cDNA; 1029 BP.
AC V45444;
DT 02-FEB-1999 (first entry)
DE Human chemokine ZSIG-35 cDNA.
KW ZSIG-35; beta-chemokine; human; ligand; lymphocyte migration;
KW inflammation; ischaemia; reperfusion injury; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 102. .554
FT /*tag= a
FT sig_peptide 102. .161
FT /*tag= b
FT mat_peptide 162. .551
FT /*tag= c
PN WO9844117-A1.
PD 08-OCT-1998.
PF 27-MAR-1998; U06115.
PR 09-MAY-1997; US-046083.
PR 28-MAR-1997; US-042862.
PA (ZYMO) ZYMOGENETICS INC.
PI Sheppard PO;
DR WPI; 98-557114/47.
DR P-PSDB; W30565.
PT New human chemokine ZSIG-35 - used for, e.g. treating inflammatory
PT disease, lymphocyte migration and ischaemia/reperfusion injury
PS Claim 15; Page 80-81; 105pp; English.
CC This nucleotide sequence includes an open reading frame for a novel
CC human chemokine (see W30565), designated ZSIG-35, that has homology
CC to members of the beta-chemokine family, in particular to murine
CC macrophage inflammatory protein. ZSIG-35 polynucleotides were
CC initially identified by querying an expressed sequence tag
CC database. The gene was mapped to chromosone 19p13.3. The
CC invention provides ZSIG-35 polynucleotides and polypeptides, an
CC expression vector, a cultured cell carrying the vector, and an
CC oligonucleotide probe or primer (see V45446). The cultured cell
CC may be used may be used for the production of protein (claimed).
CC ZSIG-35 may be used in therapeutics (claimed) for the regulation of
CC acute and chronic inflammatory disease conditions, lymphocyte
CC migration and ischaemia/reperfusion injury. The probe can be used
CC to detect a genetic abnormality in a patient (claimed) by comparing
CC the hybridisation of a genetic sample to the probe to produce a
CC first reaction product, and then comparing this product to a
CC control.
SQ Sequence 1029 BP; 241 A; 303 C; 261 G; 224 T;

Query Match 99.3%; Score 984.8; DB 1; Length 1029;
Best Local Similarity 99.8%; Pred. No. 6.5e-265;
Matches 986; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       5 GCTTGGCCTACAGCCCGGCGGGCATCAGCTCCCTTGACCCAGTGGATATCGGTGGCCCCG 64
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCTTGGCCTACAGCCCGGCGGGCATCAGCTCCCTTGACCCAGTGGATATCGGTGGCCCCG 60

Qy      65 TTATTCGTCCAGGTGCCCAGGGAGGAGGACCCGCCTGCAGCATGAACCTGTGGCTCCTGG 124
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TTATTCGTCCAGGTGCCCAGGGAGGAGGACCCGCCTGCAGCATGAACCTGTGGCTCCTGG 120

Qy     125 CCTGCCTGGTGGCCGGCTTCCTGGGAGCCTGGGCCCCCGCTGTCCACGCCCAAGGTGTCT 184
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     121 CCTGCCTGGTGGCCGGCTTCCTGGGAGCCTGGGCCCCCGCTGTCCACACCCAAGGTGTCT 180

Qy     185 TTGAGGACTGCTGCCTGGCCTACCACTACCCCATTGGGTGGGCTGTGCTCCGGCGCGCCT 244
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    181 TTGAGGACTGCTGCCTGGCCTACCACTACCCCATTGGGTGGGCTGTGCTCCGGCGCGCCT 240

Qy    245 GGACTTACCGGATCCAGGAGGTGAGCGGGAGCTGCAATCTGCCTGCTGCGATATTCTACC 304
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GGACTTACCGGATCCAGGAGGTGAGCGGGAGCTGCAATCTGCCTGCTGCGATATTCTACC 300

Qy    305 TCCCCAAGAGACACAGGAAGGTGTGTGGGAACCCCAAAAGCAGGGAGGTGCAGAGAGCCA 364
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TCCCCAAGAGACACAGGAAGGTGTGTGGGAACCCCAAAAGCAGGGAGGTGCAGAGAGCCA 360

Qy    365 TGAAGCTCCTGGATGCTCGAAATAAGGTTTTTGCAAAGCTCCGCCACAACACGCAGACCT 424
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGAAGCTCCTGGATGCTCGAAATAAGGTTTTTGCAAAGCTCCGCCACAACACGCAGACCT 420

Qy    425 TCCAAGCAGGCCCTCATGCTGTAAAGAAGTTGAGTTCTGGAAACTCCAAGTTATCATCAT 484
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    421 TCCAAGCAGGCCCTCATGCTGTAAAGAAGTTGAGTTCTGGAAACTCCAAGTTATCATCGT 480

Qy    485 CCAAGTTTAGCAATCCCATCAGCAGCAGCAAGAGGAATGTCTCCCTCCTGATATCAGCTA 544
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CCAAGTTTAGCAATCCCATCAGCAGCAGCAAGAGGAATGTCTCCCTCCTGATATCAGCTA 540

Qy    545 ATTCAGGACTGTGAGCCGGCTCATTTCTGGGCTCCATCGGCACAGGAGGGGCCGGATCTT 604
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ATTCAGGACTGTGAGCCGGCTCATTTCTGGGCTCCATCGGCACAGGAGGGGCCGGATCTT 600

Qy    605 TCTCCGATAAAACCGTCGCCCTACAGACCCAGCTGTCCCCACGCCTCTGTCTTTTGGGTC 664
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TCTCCGATAAAACCGTCGCCCTACAGACCCAGCTGTCCCCACGCCTCTGTCTTTTGGGTC 660

Qy    665 AAGTCTTAATCCCTGCACCTGAGTTGGTCCTCCCTCTGCACCCCCACCACCTCCTGCCCG 724
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 AAGTCTTAATCCCTGCACCTGAGTTGGTCCTCCCTCTGCACCCCCACCACCTCCTGCCCG 720

Qy    725 TCTGGCAACTGGAAAGAGGGAGTTGGCCTGATTTTAAGCCTTTTGCCGCTCCGGGGACCA 784
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 TCTGGCAACTGGAAAGAGGGAGTTGGCCTGATTTTAAGCCTTTTGCCGCTCCGGGGACCA 780

Qy    785 GCAGCAATCCTGGGCAGCCAGTGGCTCTTGTAGAGAAGACTTAGGATACCTCTCTCACTT 844
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GCAGCAATCCTGGGCAGCCAGTGGCTCTTGTAGAGAAGACTTAGGATACCTCTCTCACTT 840

Qy    845 TCTGTTTCTTGCCGTCCACCCCGGGCCATGCCAGTGTGTCCCTCTGGGTCCCTCCAAAAC 904
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 TCTGTTTCTTGCCGTCCACCCCGGGCCATGCCAGTGTGTCCCTCTGGGTCCCTCCAAAAC 900

Qy    905 TCTGGTCAGTTCAAGGATGCCCCTCCCAGGCTATGCTTTTCTATAACTTTTAAATAAACC 964
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TCTGGTCAGTTCAAGGATGCCCCTCCCAGGCTATGCTTTTCTATAACTTTTAAATAAACC 960

Qy    965 TTGGGGGTTGATGGAGTCAAAAAAAAAA 992
          |||||||||||||||||||||||||||
Db    961 TTGGGGGTTGATGGAGTCAAAAAAAAAA 988


RESULT   2
V04481
ID   V04481 standard; cDNA; 989 BP.
AC   V04481;
DT   22-JUN-1998  (first entry)
DE   Human chemokine beta-15 gene.
KW   Chemokine beta-15; CB15; human; thymus; thymocyte; inflammation;
KW   autoimmune disease; immunosuppression; tumour; cancers;
KW   hypoactivity; hyperactivity; atrophy; arthritis; leukaemia;
KW   lymphomas; sepsis; wound healing; myelosuppression; diagnosis;
KW   therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             88. .537
FT                   /*tag=  a
FT   sig_peptide     88. .147
FT                   /*tag=  b
FT   mat_peptide     148. .534
FT                   /*tag=  c
PN   WO9748807-A1.
PD   24-DEC-1997.
PF   17-JUN-1996; U10561.
PR   17-JUN-1996; WO-U10561.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Kreider BL, Rosen CA, Wei Y;
DR   WPI; 98-063145/06.
DR   P-PSDB; W41938.
PT   New human chemokine beta-15 gene - used to develop products for the
PT   diagnosis and treatment of thymus-related disorders, e.g.
PT   auto-immunity, immunosuppression or inflammation
PS   Claim 2; Page 58-59; 79pp; English.
CC   This polynucleotide includes a coding region for human chemokine
CC   beta-15 (CB15) (see W41938), a novel CC chemokine that modulates
CC   early thymocyte proliferation and differentiation, that mediates
CC   the differentiation of intrathymic T cell precursors into mature
CC   T-lymphocytes, and that modulates colony formation of bone marrow
CC   progenitor cells.  The nucleotide sequence was obtained by
CC   sequencing cDNA clone HTSEX82 (ATCC 97519), which was discovered in
CC   a cDNA library derived from human thymus tissue; CR15 is expressed
CC   only in tissue of the thymus.  Assay of CB15 gene expression can be
CC   used in methods for the diagnosis of disorders of the thymus such
CC   as tumours, cancers, hypoactivity, hyperactivity, atrophy,
CC   enlargement of the thymus, or disorders involving autoimmunity,
CC   arthritis, leukaemias, lymphomas, immunosuppression, sepsis, wound
CC   healing, acute and chronic inflammation, cell mediated immunity,
CC   humor immunity, or TH1/TH2 imbalance.  The isolated polynucleotide
CC   can be used for the recombinant production of CB15 useful for the
CC   treatment of individuals in need of increased levels of CB15 e.g.
CC   for myelosuppression.
SQ   Sequence   989 BP;    219 A;    297 C;    253 G;    220 T;


  Query Match            94.7%;  Score 939.6;  DB 1;  Length 989;
  Best Local Similarity  99.2%;  Pred. No. 2.4e-252;
  Matches  966;  Conservative    0;  Mismatches    4;  Indels    4;  Gaps    2;

Qy     19 CCGGCGGGCATCAGCTCCCTTGACCCAGTGGATATCGGTGGCCCCGTTATTCGTCCAGGT 78
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CCGGCGGGCATCAGCTCCCTTGACCCAGTGGATATCGGTGGCCCCGTTATTCGTCCAGGT 60

Qy     79 GCCCAGGGAGGAGGACCCGCCTGCAGCATGAACCTGTGGCTCCTGGCCTGCCTGGTGGCC 138
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCCCAGGGAGGAGGACCCGCCTGCAGCATGAACCTGTGGCTCCTGGCCTGCCTGGTGGCC 120

Qy    139 GGCTTCCTGGGAGCCTGGGCCCCCGCTGTCCACGCCCAAGGTGTCTTTGAGGACTGCTGC 198
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    121 GGCTTCCTGGGAGCCTGGGCCCCCGCTGTCCACACCCAAGGTGTCTTTGAGGACTGCTGC 180

Qy    199 CTGGCCTACCACTACCCCATTGGGTGGGCTGTGCTCCGGCGCGCCTGGACTTACCGGATC 258
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTGGCCTACCACTACCCCATTGGGTGGGCTGTGCTCCGGCGCGCCTGGACTTACCGGATC 240

Qy    259 CAGGAGGTGAGCGGGAGCTGCAATCTGCCTGCTGCGATATTCTACCTCCCCAAGAGACAC 318
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CAGGAGGTGAGCGGGAGCTGCAATCTGCCTGCTGCGATATTCTACCTCCCCAAGAGACAC 300

Qy    319 AGGAAGGTGTGTGGGAACCCCAAAAGCAGGGAGGTGCAGAGAGCCATGAAGCTCCTGGAT 378
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGGAAGGTGTGTGGGAACCCCAAAAGCAGGGAGGTGCAGAGAGCCATGAAGCTCCTGGAT 360

Qy    379 GCTCGAAATAAGGTTTTTGCAAAGCTCCGCCACAACACGCAGACCTTCCAAGCAGGCCCT 438
          |||||||||||||||||||||||||||| ||||||||||||||||||||   |||||||
Db    361 GCTCGAAATAAGGTTTTTGCAAAGCTCCACCACAACACGCAGACCTTCCA---AGGCCCT 417

Qy    439 CATGCTGTAAAGAAGTTGAGTTCTGGAAACTCCAAGTTATCATCATCCAAGTTTAGCAAT 498
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    418 CATGCTGTAAAGAAGTTGAGTTCTGGAAACTCCAAGTTATCATCGTCCAAGTTTAGCAAT 477

Qy    499 CCCATCAGCAGCAGCAAGAGGAATGTCTCCCTCCTGATATCAGCTAATTCAGGACTGTGA 558
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    478 CCCATCAGCAGCAGCAAGAGGAATGTCTCCCTCCTGATATCAGCTAATTCAGGACTGTGA 537
```

```
Qy    559 GCCGGCTCATTTCTGGGCTCCATCGGCACAGGAGGGGCCGGATCTTTCTCCGATAAAACC 618
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db    538 GCCGGCTCATTTCTGGGCTCCATCGGCACAGGA-GGGCCGGATCTTTCTCCGATAAAACC 596

Qy    619 GTCGCCCTACAGACCCAGCTGTCCCCACGCCTCTGTCTTTTGGGTCAAGTCTTAATCCCT 678
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    597 GTCGCCCTACAGACCCAGCTGTCCCCACGCCTCTGTCTTTTGGGTCAAGTCTTAATCCCT 656

Qy    679 GCACCTGAGTTGGTCCTCCCTCTGCACCCCCACCACCTCCTGCCCGTCTGGCAACTGGAA 738
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    657 GCACCTGAGTTGGTCCTCCCTCTGCACCCCCACCACCTCCTGCCCGTCTGGCAACTGGAA 716

Qy    739 AGAGGGAGTTGGCCTGATTTTAAGCCTTTTGCCGCTCCGGGGACCAGCAGCAATCCTGGG 798
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    717 AGAGGGAGTTGGCCTGATTTTAAGCCTTTTGCCGCTCCGGGGACCAGCAGCAATCCTGGG 776

Qy    799 CAGCCAGTGGCTCTTGTAGAGAAGACTTAGGATACCTCTCTCACTTTCTGTTTCTTGCCG 858
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    777 CAGCCAGTGGCTCTTGTAGAGAAGACTTAGGATACCTCTCTCACTTTCTGTTTCTTGCCG 836

Qy    859 TCCACCCCGGGCCATGCCAGTGTGTCCCTCTGGGTCCCTCCAAAACTCTGGTCAGTTCAA 918
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    837 TCCACCCCGGGCCATGCCAGTGTGTCCCTCTGGGTCCCTCCAAAACTCTGGTCAGTTCAA 896

Qy    919 GGATGCCCCTCCCAGGCTATGCTTTTCTATAACTTTTAAATAAACCTTGGGGGTTGATGG 978
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    897 GGATGCCCCTCCCAGGCTATGCTTTTCTATAACTTTTAAATAAACCTTGGGGGGTGATGG 956

Qy    979 AGTCAAAAAAAAAA 992
          ||||||||||||||
Db    957 AGTCAAAAAAAAAA 970
```

```
RESULT   3
V15415
ID   V15415 standard; cDNA; 1012 BP.
AC   V15415;
DT   11-JUN-1998  (first entry)
DE   Human thymus expressed chemokine encoding cDNA.
KW   Human; thymus expressed chemokine; TECK; MIP-3alpha; MIP-3beta;
KW   receptor; dendritic cell; macrophage; inflammation; asthma; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             117. .569
FT                   /*tag=  a
FT                   /product= "thymus expressed chemokine"
FT   sig_peptide     117. .185
FT                   /*tag=  b
FT   mat_peptide     186. .566
FT                   /*tag=  c
FT                   /product= "TECK"
PN   WO9801557-A2.
PD   15-JAN-1998.
PF   02-JUL-1997; U10819.
PR   04-JUN-1997; US-048593.
PR   05-JUL-1996; US-675814.
PR   11-OCT-1996; US-028329.
PA   (SCHE ) SCHERING CORP.
PI   Gish KC, Schall TJ, Vicari A, Wang W, Zlotnik A;
DR   WPI; 98-101054/09.
DR   P-PSDB; W44397.
PT   Novel chemokines, e.g. thymus expressed chemokine - used for
PT   treating inflammatory conditions including asthma.
PS   Claim 5; Page 87-88; 202pp; English.
CC   The present sequence encodes a thymus expressed chemokine (TECK).
CC   Antibodies which bind to the protein can be used in detecting or
CC   diagnosing various immunological conditions related to expression
CC   of the protein. The nucleic acid can be used for screening and
CC   isolating DNA clones for the chemokines, especially from other
CC   species. The chemokine can be used in the treatment of conditions
CC   associated with abnormal physiology or development, including
CC   inflammatory conditions such as asthma.
```

```
SQ   Sequence   1012 BP;    225 A;    307 C;    254 G;    226 T;

  Query Match             92.2%;  Score 914.4;  DB 1;  Length 1012;
  Best Local Similarity   97.1%;  Pred. No. 2.5e-245;
  Matches  963;  Conservative    0;  Mismatches   26;  Indels    3;  Gaps    3;

Qy      1 GACAGCTTGGCCTACAGCCCGGCGGGCATCAGCTCCCTTGACCCAGTGGATATCGGTGGC 60
          | | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12 GTCCGCTTGGCCTACAGCCCGGCGGGCATCAGCTCCCTTGACCCAGTGGATATCGGTGGC 71

Qy     61 CCCGTTATTCGTCCAGGTGCCCAGGGAGGAGGACCCGCCTGCAGCATGAACCTGTGGCTC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     72 CCCGTTATTCGTCCAGGTGCCCAGGGAGGAGGACCCGCCTGCAGCATGAACCTGTGGCTC 131

Qy    121 CTGGCCTGCCTGGTGGCCGGCTTCCTGGGAGCCTGGGCCCCCGCTGTCCACGCCCAAGGT 180
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    132 CTGGCCTGCCTGGTGGCCGGCTTCCTGGGAGCCTGGGCCCCCGCTGTCCACACCCAAGGT 191

Qy    181 GTCTTTGAGGACTGCTGCCTGGCCTACCACTACCCCATTGGGTGGGCTGTGCTCCGGCGC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    192 GTCTTTGAGGACTGCTGCCTGGCCTACCACTACCCCATTGGGTGGGCTGTGCTCCGGCGC 251

Qy    241 GCCTGGACTTACCGGATCCAGGAGGTGAGCGGGAGCTGCAATCTGCCTGCTGCGATATTC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    252 GCCTGGACTTACCGGATCCAGGAGGTGAGCGGGAGCTGCAATCTGCCTGCTGCGATATTC 311

Qy    301 TACCTCCCCAAGAGACACAGGAAGGTGTGTGGGAACCCCAAAAGCAGGGAGGTGCAGAGA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    312 TACCTCCCCAAGAGACACAGGAAGGTGTGTGGGAACCCCAAAAGCAGGGAGGTGCAGAGA 371

Qy    361 GCCATGAAGCTCCTGGATGCTCGAAATAAGGTTTTTGCAAAGCTCCGCCACAACACGCAG 420
          |||||||||||||||||||||||||||||||||||||||||||||||| |||||||| ||||
Db    372 GCCATGAAGCTCCTGGATGCTCGAAATAAGGTTTTTGCAAAGCTCCACCACAACATGCAG 431

Qy    421 ACCTTCCAAGCAGGCCCTCATGCTGTAAAGAAGTTGAGTTCTGGAAACTCCAAGTTATCA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    432 ACCTTCCAAGCAGGCCCTCATGCTGTAAAGAAGTTGAGTTCTGGAAACTCCAAGTTATCA 491

Qy    481 TCATCCAAGTTTAGCAATCCCATCAGCAGCAGCAAGAGGAATGTCTCCCTCCTGATATCA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    492 TCATCCAAGTTTAGCAATCCCATCAGCAGCAGCAAGAGGAATGTCTCCCTCCTGATATCA 551

Qy    541 GCTAATTCAGGACTGTGAGCCGGCTCATTTCTGGGCTCCATCGGCACAGGAGGGGCCGGA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    552 GCTAATTCAGGACTGTGAGCCGGCTCATTTCTGGGCTCCATCGGCACAGGAGGGGCCGGA 611

Qy    601 TCTTTCTCCGATAAAACCGTCGCCCTACAGACCCAGCTGTCCCCACGCCTCTGTCTTTTG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    612 TCTTTCTCCGATAAAACCGTCGCCCTACAGACCCAGCTGTCCCCACGCCTCTGTCTTTTG 671

Qy    661 GGTCAAGTCTTAATCCCTGCACCTGAGTTGGTCCTCCCTCTGCACCCCCACCACCTCCTG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    672 GGTCAAGTCTTAATCCCTGCACCTGAGTTGGTCCTCCCTCTGCACCCCCACCACCTCCTG 731

Qy    721 CCCGTCTGGCAACTGGAAAGAGGGAGTTGGCCTGATTTTAAGCCTTTTGCCGCTCCGGGG 780
          |||||||||||||||||| |||||||||||||||| |||||||||||||||||||||||
Db    732 CCCGTCTGGCAACTGGAAAGAAGGAGTTGGCCTGATTTTAA-CCTTTTGCCGCTCCGGGG 790

Qy    781 ACCAGCAGCAATCCTGGGCAGCCAGTGGCTCTTGTAGAGAAGACTTAGGATACCTCTCTC 840
          | ||||| |||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    791 AACAGCA-CAATCCTGGGCAGCCAGTGGCTCTTGTAGAGAAAACTTAGGATACCTCTCTC 849

Qy    841 ACTTTCTGTTTCTTGCCGTCCACCCCGGGCCATGCCAGTGTGTCCCTCTGGGTCCCTCCA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||    ||  |||||||
Db    850 ACTTTCTGTTTCTTGCCGTCCACCCCGGGCCATGCCAGTGTGTCCCTCTGGGTCCCCTCCA 909

Qy    901 AAACTCTGGTCAGTTCAAGGATGCCCCTCCCAGGCTATGCTTTTCTATAACTTTTAAATA 960
          ||| |||||||| |||||||| |||  || ||||||||||||||||||||||||||||||
Db    910 AAAATCTGGTCA-TTCAAGGATCCCCTCCCAAGGCTATGCTTTTCTATAACTTTTAAATA 968

Qy    961 AACCTTGGGGGTTGATGGAGTCAAAAAAAAAA 992
          |||||||||||| |||  |    |||||||||
```

```
Db    969 AACCTTGGGGGGTGAATGGAATAAAAAAAAAA 1000


RESULT   4
V45445
ID   V45445 standard; DNA; 450 BP.
AC   V45445;
DT   02-FEB-1999  (first entry)
DE   Human chemokine ZSIG-35 DNA probe.
KW   ZSIG-35; beta-chemokine; human; ligand; lymphocyte migration;
KW   inflammation; ischaemia; reperfusion injury; probe; ss.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9844117-A1.
PD   08-OCT-1998.
PF   27-MAR-1998; U06115.
PR   09-MAY-1997; US-046083.
PR   28-MAR-1997; US-042862.
PA   (ZYMO ) ZYMOGENETICS INC.
PI   Sheppard PO;
DR   WPI; 98-557114/47.
PT   New human chemokine ZSIG-35 - used for, e.g. treating inflammatory
PT   disease, lymphocyte migration and ischaemia/reperfusion injury
PS   Claim 14; Page 88; 105pp; English.
CC   This degenerate polynucleotide encompasses all DNAs that encode
CC   the novel human chemokine ZSIG-35 (see also V45444 and W30565) of
CC   the invention.  Probes and primers comprising at least 14
CC   contiguous nucleotides of this polynucleotide, or its complementary
CC   sequence, can be used to detect a genetic abnormality in a patient
CC   (claimed) by comparing the hybridisation of a genetic sample to
CC   the probe to produce a first reaction product, and then comparing
CC   this product to a control.  ZSIG-35 may be used in therapeutics for
CC   the regulation of acute and chronic inflammatory disease conditions,
CC   lymphocyte migration and ischaemia/reperfusion injury.
SQ   Sequence   450 BP;    70 A;    44 C;    74 G;    59 T;


  Query Match             31.4%;  Score 311.8;  DB 1;  Length 450;
  Best Local Similarity   54.8%;  Pred. No. 9e-78;
  Matches  246;  Conservative  114;  Mismatches   89;  Indels    0;  Gaps    0;

Qy    106 ATGAACCTGTGGCTCCTGGCCTGCCTGGTGGCCGGCTTCCTGGGAGCCTGGGCCCCCGCT 165
          |||||::| |||:| :| || ||::| || || || ||::| || || ||||| || ||
Db      1 ATGAAYYTNTGGYTNYTNGCNTGYYTNGTNGCNGGNTTYYTNGGNGCNTGGGCNCCNGCN 60

Qy    166 GTCCACGCCCAAGGTGTCTTTGAGGACTGCTGCCTGGCCTACCACTACCCCATTGGGTGG 225
          || ||: | ||:|| || ||:||:||:||:||:|| || ||:||:||:|| ||:|| |||
Db     61 GTNCAYACNCARGGNGTNTTYGARGAYTGYTGYYTNGCNTAYCAYTAYCCNATHGGNTGG 120

Qy    226 GCTGTGCTCCGGCGCGCCTGGACTTACCGGATCCAGGAGGTGAGCGGGAGCTGCAATCTG 285
          || || :| :| :| || ||||| ||::| ||:||:||:|| :: || :: ||:||::|
Db    121 GCNGTNYTNMGNMGNGCNTGGACNTAYMGNATHCARGARGTNWSNGGNWSNTGYAAYYTN 180

Qy    286 CCTGCTGCGATATTCTACCTCCCCAAGAGACACAGGAAGGTGTGTGGGAACCCCAAAAGC 345
          || || || ||:||:||::| || ||::| ||::| ||:|| ||:|| ||:|| ||:::
Db    181 CCNGCNGCNATHTTYTAYYTNCCNAARMGNCAYMGNAARGTNTGYGGNAAYCCNAARWSN 240

Qy    346 AGGGAGGTGCAGAGAGCCATGAAGCTCCTGGATGCTCGAAATAAGGTTTTTGCAAAGCTC 405
          :| ||:|| ||::| || |||||::| :| ||:|| :| ||:||:|| ||:|| ||:::|
Db    241 MGNGARGTNCARMGNGCNATGAARYTNYTNGAYGCNMGNAAYAARGTNTTYGCNAARYTN 300

Qy    406 CGCCACAACACGCAGACCTTCCAAGCAGGCCCTCATGCTGTAAAGAAGTTGAGTTCTGGA 465
          :| ||:||:|| ||:|| ||:||:|| || || ||:|| || ||:||::| :: :: ||
Db    301 MGNCAYAAYACNCARACNTTYCARGCNGGNCCNCAYGCNGTNAARAARYTNWSNWSNGGN 360

Qy    466 AACTCCAAGTTATCATCATCCAAGTTTAGCAATCCCATCAGCAGCAGCAAGAGGAATGTC 525
          ||::: ||::| :: :: :: ||:||::: ||:|| ||::: :: :: ||::| ||:||
Db    361 AAYWSNAARYTNWSNWSNWSNAARTTYWSNAAYCCNATHWSNWSNWSNAARMGNAAYGTN 420

Qy    526 TCCCTCCTGATATCAGCTAATTCAGGACT 554
          :: :| :| ||::: || ||::: || :|
Db    421 WSNYTNYTNATHWSNGCNAAYWSNGGNYT 449
```

```
RESULT   5
V15414
ID   V15414 standard; cDNA; 1034 BP.
AC   V15414;
DT   11-JUN-1998  (first entry)
DE   Mouse thymus expressed chemokine encoding cDNA.
KW   Mouse; thymus expressed chemokine; TECK; MIP-3alpha; MIP-3beta;
KW   receptor; dendritic cell; macrophage; inflammation; asthma; ss.
OS   Mus sp.
FH   Key             Location/Qualifiers
FT   CDS             94. .528
FT                   /*tag=  a
FT                   /product= "thymus expressed chemokine"
FT   sig_peptide     94. .162
FT                   /*tag=  b
FT                   /note= "putative"
FT   mat_peptide     163. .525
FT                   /*tag=  c
FT                   /product= "TECK"
PN   WO9801557-A2.
PD   15-JAN-1998.
PF   02-JUL-1997; U10819.
PR   04-JUN-1997; US-048593.
PR   05-JUL-1996; US-675814.
PR   11-OCT-1996; US-028329.
PA   (SCHE ) SCHERING CORP.
PI   Gish KC, Schall TJ, Vicari A, Wang W, Zlotnik A;
DR   WPI; 98-101054/09.
DR   P-PSDB; W44396.
PT   Novel chemokines, e.g. thymus expressed chemokine - used for
PT   treating inflammatory conditions including asthma.
PS   Claim 5; Page 85-86; 202pp; English.
CC   The present sequence encodes a thymus expressed chemokine (TECK).
CC   Antibodies which bind to the protein can be used in detecting or
CC   diagnosing various immunological conditions related to expression
CC   of the protein. The nucleic acid can be used for screening and
CC   isolating DNA clones for the chemokines, especially from other
CC   species. The chemokine can be used in the treatment of conditions
CC   associated with abnormal physiology or development, including
CC   inflammatory conditions such as asthma.
SQ   Sequence   1034 BP;    320 A;    212 C;    272 G;    230 T;


  Query Match             16.6%;  Score 165;  DB 1;  Length 1034;
  Best Local Similarity   67.4%;  Pred. No. 8.8e-37;
  Matches  250;  Conservative    0;  Mismatches  115;  Indels    6;  Gaps    1;

Qy     22 GCGGGCATCAGCTCCCTTGACCCAGTGGATATCGGTGGCCCCGTTATTCGTCCAGGTGCC 81
          || |||| |||||| |  ||||      |  || |||||||||| |||  | | |||||  |
Db     10 GCAGGCACCAGCTCTCAGGACCAGAAAGGCATTGGTGGCCCCCTTAAACCTTCAGGTATC 69

Qy     82 CAGGGAGGAGGACCCGCCTGCAGCATGAACCTGTGGCTCCTGGCCTGCCTGGTGGCCGGC 141
           | ||||||  |   ||| ||  ||||| ||||||||  | ||||||||||| ||| |
Db     70 TGGAGAGGAGATCTAACCTTCACTATGAAACTGTGGCTTTTTGCCTGCCTGGTTGCCTGT 129

Qy    142 TTCCTGGGAGCCTGGGCCCCCGCTGTCCACGCCCAAGGTGTCTTTGAGGACTGCTGCCTG 201
          ||  | || ||||||   || | |||||| |||||||||| |||||| ||||||||||||
Db    130 TTTGTTGGGGCCTGGATGCCGGTTGTCCATGCCCAAGGTGCCTTTGAAGACTGCTGCCTG 189

Qy    202 GCCTACCACTACCCCATTGGGTGGGCTGTGCTCCGGCGCGCCTGGACTTACCGGATCCAG 261
          |  |||||  ||   ||    |||  ||| |||||||  ||  ||| ||| |     |||
Db    190 GGTTACCAGCACAGGATCAAATGGAATGTTCTCCGGCATGCTAGGAATTATCACCAGCAG 249

Qy    262 GAGGTGAGCGGGAGCTGCAATCTGCCTGCTGCGATATTCTACCTCCCCAAGAGACACAGG 321
          || ||||| || |||||||| || | ||||| || ||||||| ||| | ||| |
Db    250 GAAGTGAGTGGAAGCTGCAACCTACGTGCTGTGAGATTCTACTTCCGCCAGAAA------ 304

Qy    322 AAGGTGTGTGGGAACCCCAAAAGCAGGGAGGTGCAGAGAGCCATGAAGCTCCTGGATGCT 381
            ||||||||||| ||  |   || | | ||| |||| || || |   || ||   |||
Db    304 GTAGTGTGTGGGAATCCAGAGGACATGAATGTGAAGAGGGCGATAAGAATCTTGACAGCT 363

Qy    382 CGAAATAAGGT 392
```

```
          | || | | |
Db    364 AGGAAAAGGCT 374


RESULT   6
V63290
ID   V63290 standard; DNA; 285 BP.
AC   V63290;
DT   21-JAN-1999  (first entry)
DE   Codon optimised human chemokine beta-6 DNA for expression in E. coli.
KW   Human; chemokine beta-6; Ckb-6; agonist; antagonist;
KW   basophil activation; eosinophil; histamine release; myeloprotection;
KW   chemotherapy; ss.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9844118-A1.
PD   08-OCT-1998.
PF   31-MAR-1998; U06401.
PR   19-DEC-1997; US-995156.
PR   31-MAR-1997; US-042269.
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (KOCH-) KOCHER INST THEODOR.
PI   Baggiolini M, Kreider BL, Olsen HS, Ruben SM;
DR   WPI; 98-557115/47.
PT   New isolated chemokine beta-6 - used to develop products for
PT   treating e.g. autoimmune disorders, neurodegenerative disorders,
PT   inflammation, allergy, infectious diseases or cancers
PS   Example 13; Page 116; 163pp; English.
CC   The present sequence represents a codon optimised human chemokine
CC   beta-6 (Ckb-6) DNA sequence for expression in Escherichia coli.
CC   The Ckb-6 polypeptide can act as a Ckb-6 agonist or antagonist. The
CC   protein can be used for activating or mobilising basophils or
CC   eosinophils, for stimulating histamine release or for
CC   myeloprotection for a patient undergoing chemotherapy.
SQ   Sequence   285 BP;    70 A;    85 C;    71 G;    59 T;


  Query Match             4.1%;  Score 40.2;  DB 1;  Length 285;
  Best Local Similarity   56.4%;  Pred. No. 0.029;
  Matches   75;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy    258 CCAGGAGGTGAGCGGGAGCTGCAATCTGCCTGCTGCGATATTCTACCTCCCCAAGAGACA 317
          ||||  |   ||| | |||  |   |||   ||||   |  ||| |  | |||| |
Db     75 CCAGCTGTCCAGCCGCAGCACCTGCCTGAAAGCTGGCGTGATCTTCACCACCAAAAAGGG 134

Qy    318 CAGGAAGGTGTGTGGGAACCCCAAAAGCAGGGAGGTGCAGAGAGCCATGAAGCTCCTGGA 377
          |  | || | |||||  |||| |||     |  ||| ||| |   ||||||   ||||||
Db    135 CCAGCAGTTCTGTGGCGACCCGAAACAAGAGTGGGTCCAGCGTTACATGAAAAACCTGGA 194

Qy    378 TGCTCGAAATAAG 390
           ||   | | |||
Db    195 CGCCAAACAGAAG 207


RESULT   7
V63305
ID   V63305 standard; DNA; 4256 BP.
AC   V63305;
DT   21-JAN-1999  (first entry)
DE   Nucleotide sequence of pHE4-5 expressaion vector used to clone Ckb-6.
KW   Human; chemokine beta-6; Ckb-6; agonist; antagonist;
KW   basophil activation; eosinophil; histamine release; myeloprotection;
KW   chemotherapy; ss.
OS   Synthetic.
PN   WO9844118-A1.
PD   08-OCT-1998.
PF   31-MAR-1998; U06401.
PR   19-DEC-1997; US-995156.
PR   31-MAR-1997; US-042269.
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (KOCH-) KOCHER INST THEODOR.
PI   Baggiolini M, Kreider BL, Olsen HS, Ruben SM;
DR   WPI; 98-557115/47.
PT   New isolated chemokine beta-6 - used to develop products for
PT   treating e.g. autoimmune disorders, neurodegenerative disorders,
PT   inflammation, allergy, infectious diseases or cancers
PS   Disclosure; Pages 121-123; 163pp; English.
CC   The present sequence represents expression vector pHEA4-5, which
CC   was used to clone DNA encoding a human chemokine beta-6 (Ckb-6)
CC   polypeptide. The Ckb-6 protein can act as a Ckb-6 agonist or
CC   antagonist. The protein can be used for activating or mobilising
CC   basophils or eosinophils, for stimulating histamine release or for
CC   myeloprotection for a patient undergoing chemotherapy.
SQ   Sequence   4256 BP;    968 A;    1201 C;    1165 G;    922 T;


  Query Match             4.1%;  Score 40.2;  DB 1;  Length 4256;
  Best Local Similarity   56.4%;  Pred. No. 0.082;
  Matches   75;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy    258 CCAGGAGGTGAGCGGGAGCTGCAATCTGCCTGCTGCGATATTCTACCTCCCCAAGAGACA 317
          ||||  |   ||| | |||  |   |||   ||||   |  ||| |  | |||| |
Db    184 CCAGCTGTCCAGCCGCAGCACCTGCCTGAAAGCTGGCGTGATCTTCACCACCAAAAAGGG 243

Qy    318 CAGGAAGGTGTGTGGGAACCCCAAAAGCAGGGAGGTGCAGAGAGCCATGAAGCTCCTGGA 377
          |  | || | |||||  |||| |||     |  ||| ||| |   ||||||   ||||||
Db    244 CCAGCAGTTCTGTGGCGACCCGAAACAAGAGTGGGTCCAGCGTTACATGAAAAACCTGGA 303

Qy    378 TGCTCGAAATAAG 390
           ||   | | |||
Db    304 CGCCAAACAGAAG 316


RESULT   8
V63289
ID   V63289 standard; DNA; 360 BP.
AC   V63289;
DT   21-JAN-1999  (first entry)
DE   DNA encoding human chemokine beta-6.
KW   Human; chemokine beta-6; Ckb-6; agonist; antagonist;
KW   basophil activation; eosinophil; histamine release; myeloprotection;
KW   chemotherapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             1. .360
FT                   /*tag=  a
FT   sig_peptide     1. .78
FT                   /*tag=  b
FT   mat_peptide     79. .357
FT                   /*tag=  c
PN   WO9844118-A1.
PD   08-OCT-1998.
PF   31-MAR-1998; U06401.
PR   19-DEC-1997; US-995156.
PR   31-MAR-1997; US-042269.
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (KOCH-) KOCHER INST THEODOR.
PI   Baggiolini M, Kreider BL, Olsen HS, Ruben SM;
DR   WPI; 98-557115/47.
DR   P-PSDB; W70593.
PT   New isolated chemokine beta-6 - used to develop products for
PT   treating e.g. autoimmune disorders, neurodegenerative disorders,
PT   inflammation, allergy, infectious diseases or cancers
PS   Example 1; Fig 1; 163pp; English.
CC   The present sequence encodes a human chemokine beta-6 (Ckb-6)
CC   polypeptide which can act as a Ckb-6 agonist or antagonist. The
CC   protein can be used for activating or mobilising basophils or
CC   eosinophils, for stimulating histamine release or for
CC   myeloprotection for a patient undergoing chemotherapy.
SQ   Sequence   360 BP;    85 A;    107 C;    95 G;    73 T;


  Query Match             4.1%;  Score 40.2;  DB 1;  Length 360;
  Best Local Similarity   56.4%;  Pred. No. 0.031;
  Matches   75;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;
```

```
Qy    258 CCAGGAGGTGAGCGGGAGCTGCAATCTGCCTGCTGCGATATTCTACCTCCCCAAGAGACA 317
          ||||  |   ||| |||||      ||    || |   |  ||| |  | ||||||
Db    150 CCAGCTGTCCAGCAGGAGCACATGCCTCAAGGCAGGAGTGATCTTCACCACCAAGAAGGG 209

Qy    318 CAGGAAGGTGTGTGGGAACCCCAAAAGCAGGGAGGTGCAGAGAGCCATGAAGCTCCTGGA 377
          |  | || | |||||  |||||||      |  ||| |||||   |||||||  ||||||
Db    210 CCAGCAGTTCTGTGGCGACCCCAAGCAGGAGTGGGTCCAGAGGTACATGAAGAACCTGGA 269

Qy    378 TGCTCGAAATAAG 390
           ||     | |||
Db    270 CGCCAAGCAGAAG 282


RESULT   9
T03207
ID   T03207 standard; cDNA; 360 BP.
AC   T03207;
DT   15-JUN-1996  (first entry)
DE   Human monocyte chemotactic protein-4 cDNA.
KW   Monocyte chemotactic protein-4; MCP-4; antitumor; vulnerary;
KW   antiparasitic; antiinflammatory; antiallergic; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             1. .360
FT                   /*tag=  a
FT   signal_peptide  1. .66
FT                   /*tag=  b
FT   mat_peptide     67. .360
FT                   /*tag=  c
FT                   /product= MCP-4
PN   WO9531467-A1.
PD   23-NOV-1995.
PF   16-MAY-1994; U05384.
PR   16-MAY-1994; WO-U05384.
PR   26-MAY-1994; ZA-003695.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Li H,  Ruben S,  Sutton GG;
DR   WPI; 96-010869/01.
DR   P-PSDB; R85779.
PT   Monocyte chemotactic protein-4 and corresponding DNA - for treating
PT   tumours, wounds, parasitic infection, haematopoiesis, inflammation,
PT   allergies, etc.
PS   Claim 5; Page 29; 40pp; English.
CC   This cDNA sequence may be used to express the MCP-4 protein in a
CC   suitable host cell (e.g. yeast, fungus, bacterium, mammal, etc.).
CC   The sequence may be used to treat tumours, wounds, parasitic
CC   infections, regulation of heamatopoiesis, inflammation, rheumatoid
CC   arthritis, lung inflammation, allergies, atherosclerosis and
CC   infectious diseases such as tuberculosis.  Typically, this DNA
CC   encoding MCP-4 (ATCC 75703) was initially amplified with PCR
CC   primers and expressed in Escherichia coli m15rep4.
SQ   Sequence   360 BP;    85 A;    106 C;    96 G;    73 T;


  Query Match             3.9%;  Score 38.6;  DB 1;  Length 360;
  Best Local Similarity   55.6%;  Pred. No. 0.087;
  Matches   74;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy    258 CCAGGAGGTGAGCGGGAGCTGCAATCTGCCTGCTGCGATATTCTACCTCCCCAAGAGACA 317
          ||||  |   ||| |||||      ||    |  |   |  ||| |  | ||||||
Db    150 CCAGCTGTCCAGCAGGAGCACATGCCTCAAGGGAGGAGTGATCTTCACCACCAAGAAGGG 209

Qy    318 CAGGAAGGTGTGTGGGAACCCCAAAAGCAGGGAGGTGCAGAGAGCCATGAAGCTCCTGGA 377
          |  | || | |||||  |||||||      |  ||| |||||   |||||||  ||||||
Db    210 CCAGCAGTTCTGTGGCGACCCCAAGCAGGAGTGGGTCCAGAGGTACATGAAGAACCTGGA 269

Qy    378 TGCTCGAAATAAG 390
           ||     | |||
Db    270 CGCCAAGCAGAAG 282


RESULT  10
T47229
ID   T47229 standard; cDNA; 360 BP.
AC   T47229;
DT   09-JUL-1997  (first entry)
DE   Human monocyte chemotactic protein-4 polypeptide encoding cDNA.
KW   MCP-4; haematopoietic progenitor cell; anti-tumour; arthritis;
KW   atherosclerosis; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   signal_peptide  1. .78
FT                   /*tag=  a
FT   mat_peptide     79. .357
FT                   /*tag=  b
FT                   /product= Monocyte_chemotactic_protein-4
PN   WO9640762-A1.
PD   19-DEC-1996.
PF   07-JUN-1996; U10087.
PR   07-JUN-1995; US-479126.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Li H,  Ruben SM,  Sutton GG;
DR   WPI; 97-052228/05.
DR   P-PSDB; W07845.
PT   DNA encoding human monocyte chemotactic protein-4 - useful to
PT   increase amt., or inhibit destruction, of haematopoietic progenitor
PT   cells
PS   Claim 1; Page 51; 73pp; English.
CC   The present sequence encodes human monocyte chemotactic protein-4
CC   polypeptide (MCP-4). The polypeptide can be used in the treatment of a
CC   patient with a disease, or susceptibility to a disease, related to MCP-4
CC   under-expression. This can be diagnosed by determining a mutation in the
CC   nucleic acid sequence encoding MCP-4. The polypeptide and the
CC   polynucleotide can also be used to screen for (ant)agonists. The
CC   polypeptide can be used to increase the amount, or inhibit the
CC   destruction, of haematopoietic progenitor cells. It also inhibits
CC   degeneration of neuronal cells and can be used to promote wound healing,
CC   as an anti-tumour agent, to combat parasitic infections and to treat
CC   arthritis, lung inflammation, allergy and atherosclerosis. MCP-4
CC   shows 39% homology with MIP-1 alpha and 34% homology with MCP-1.
SQ   Sequence   360 BP;    85 A;    106 C;    96 G;    73 T;


  Query Match             3.9%;  Score 38.6;  DB 1;  Length 360;
  Best Local Similarity   55.6%;  Pred. No. 0.087;
  Matches   74;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy    258 CCAGGAGGTGAGCGGGAGCTGCAATCTGCCTGCTGCGATATTCTACCTCCCCAAGAGACA 317
          ||||  |   ||| |||||      ||    |  |   |  ||| |  | ||||||
Db    150 CCAGCTGTCCAGCAGGAGCACATGCCTCAAGGGAGGAGTGATCTTCACCACCAAGAAGGG 209

Qy    318 CAGGAAGGTGTGTGGGAACCCCAAAAGCAGGGAGGTGCAGAGAGCCATGAAGCTCCTGGA 377
          |  | || | |||||  |||||||      |  ||| |||||   |||||||  ||||||
Db    210 CCAGCAGTTCTGTGGCGACCCCAAGCAGGAGTGGGTCCAGAGGTACATGAAGAACCTGGA 269

Qy    378 TGCTCGAAATAAG 390
           ||     | |||
Db    270 CGCCAAGCAGAAG 282


RESULT  11
T94505/c
ID   T94505 standard; DNA; 1068 BP.
AC   T94505;
DT   12-FEB-1998  (first entry)
DE   Polyamide hydrolase I coding sequence.
KW   Polyamide hydrolase; polymer; nylon; recovery; production; enzyme; PCR;
KW   hydrolysis; degree of polymerisation; monomer; amplification; primer; ss.
OS   Corynebacterium sp.
FH   Key             Location/Qualifiers
FT   CDS             1. .1068
FT                   /*tag=  a
FT                   /gene= "pamI"
FT                   /product= polyamide hydrolase
PN   WO9704084-A1.
PD   06-FEB-1997.
```

PF 17-JUL-1996; F01119.
PR 18-JUL-1995; FR-008916.
PA (RHON) RHONE POULENC FIBRES & POLYMERES SA.
PI Crouzet J, Favre-Bulle O, Jourdat C, Le Coq AM, Petre D;
DR WPI; 97-132627/12.
DR P-PSDB; W29999.
PT New amidase active against polyamide, esp. of the nylon type - also
PT new DNA and transformed cells, esp. for recovering polyamide
PT monomers
PS Claim 6; Page 17-19; 31pp; French.
CC This is the nucleotide sequence encoding a polyamide hydrolase which is
CC active against a wide variety of amide (co)polymers, particularly those
CC of the nylon 6.6 type, especially for the recovery of the starting
CC materials for polyamide production. The enzyme preferably hydrolyses
CC substrates where the degrees of polymerisation (DP) are below 40,
CC resulting in product of DP below 8, preferably to monomers. The sequence
CC is isolated from strain deposited with the CNCM as I1495. The coding
CC sequence was part amplified with the primers T85989-90 in order to
CC introduce a restriction enzyme site at the 5' end of the sequence. The
CC amplified product was inserted into the plasmid pUC29 and the complete
CC gene was reconstituted by recombinant methods to generate the plasmid
CC pXL2564 in which the pamI gene is placed under control of the E. coli
CC trpE promoter and has the lambda phage CII ribosomal binding site
CC immediately preceding the coding sequence. This plasmid is transformed
CC into E. coli TG1 for production of the protein.
SQ Sequence 1068 BP; 152 A; 388 C; 382 G; 146 T;

Query Match 3.6%; Score 35.6; DB 1; Length 1068;
Best Local Similarity 48.5%; Pred. No. 0.91;
Matches 98; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

```
Qy      91 GGACCCGCCTGCAGCATGAACCTGTGGCTCCTGGCCTGCCTGGTGGCCGGCTTCCTGGGA 150
           || | ||    | |   | |||||     |     |  ||||||  | ||    || ||
Db     792 GGCCTCGGTGACGGGCGGGACCTGCTCGGCGAACGCCTCCTGGTAGTCGAACACCGGGTG 733

Qy     151 GCCTGGGCCCCCCGCTGTCCACGCCCAAGGTGTCTTTGAGGACTGCTGCCTGGCCTACCAC 210
           ||   || |||||  | | | || |  || || |    |  || |   ||  ||   ||||
Db     732 GCGCCGGACCCCGGTCTGCGCGTCGTAGTTGCCGCGCACCACCGTGCCCGCGCGGTCCAC 673

Qy     211 TACCCCATTGGGTGGGCTGTGCTCCGGCGCGCCTGGACTTACCGGATCCAGGAGGTGAGC 270
           | | ||  |   |||  |  || |     || |  |  |   |  ||  |||  ||
Db     672 GATCACACCGACCGGGTTCGGCACGACGACGGCGAGGATGCGCACGTCGCCGAGACGACG 613

Qy     271 GGGAGCTGCAATCTGCCTGCTG 292
           |   || ||   ||||| | ||
Db     612 GAACGCCGCGCCCTGCCCGGTG 591
```

RESULT 12
T94507/c
ID T94507 standard; DNA; 1068 BP.
AC T94507;
DT 12-FEB-1998 (first entry)
DE Polyamide hydrolase I coding sequence.
KW Polyamide hydrolase; Comamonas acidovorans; enzyme; hydrolysis; polymer;
KW diamine; dicarboxylic acid; nylon; recovery; monomer; Corynebacterium;
KW chromatography; primer; PCR; amplification; ss.
OS Corynebacterium sp.
FH Key Location/Qualifiers
FT CDS 1. .1068
FT /*tag= a
FT /gene= "pamI"
FT /product= polyamide hydrolase
PN WO9704083-A1.
PD 06-FEB-1997.
PF 17-JUL-1996; F01118.
PR 18-JUL-1995; FR-008917.
PA (RHON) RHONE POULENC FIBRES & POLYMERES SA.
PI Crouzet J, Faucher D, Favre-Bulle O, Jourdat C, Petre D;
PI Pierrard J, Thibaut D, Guitton C;
DR WPI; 97-132626/12.
DR P-PSDB; W36011.
PT Enzymatic hydrolysis of polyamide, esp. nylon, oligomers and new
PT enzyme - esp. from Comamonas acidovorans, DNA and transformed cells,
PT partic. for recovery of monomers
PS Example 4; Page 22; 36pp; French.
CC This is the nucleotide sequence encoding the polyamide hydrolase I from
CC a Corynebacterium sp., which is used in a novel method of enzymatic
CC hydrolysis of a wide range of (poly)amide substrates of derived from
CC diamine and dicarboxylic acid, partic. nylon 6.6, esp. for recovery of
CC polyamide monomers. Also used in the process is the polyamide
CC hydrolase II (W36010) derived from Comamonas acidovorans strain N12
CC (CNCM I-1522). The coding sequence was part amplified with the primers
CC T94508-9 in order to introduce a restriction enzyme site at the 5' end
CC of the sequence. The amplified product was inserted into the plasmid
CC pUC29 and the complete gene was reconstituted by recombinant methods to
CC generate the plasmid pXL2564 in which the pamI gene is placed under
CC control of the E. coli trpE promoter and has the lambda phage CII
CC ribosomal binding site immediately preceding the coding sequence. This
CC plasmid is transformed into E. coli TG1 for production of the protein.
SQ Sequence 1068 BP; 152 A; 388 C; 382 G; 146 T;

Query Match 3.6%; Score 35.6; DB 1; Length 1068;
Best Local Similarity 48.5%; Pred. No. 0.91;
Matches 98; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

```
Qy      91 GGACCCGCCTGCAGCATGAACCTGTGGCTCCTGGCCTGCCTGGTGGCCGGCTTCCTGGGA 150
           || | ||    | |   | |||||     |     |  ||||||  | ||    || ||
Db     792 GGCCTCGGTGACGGGCGGGACCTGCTCGGCGAACGCCTCCTGGTAGTCGAACACCGGGTG 733

Qy     151 GCCTGGGCCCCCCGCTGTCCACGCCCAAGGTGTCTTTGAGGACTGCTGCCTGGCCTACCAC 210
           ||   || |||||  | | | || |  || || |    |  || |   ||  ||   ||||
Db     732 GCGCCGGACCCCGGTCTGCGCGTCGTAGTTGCCGCGCACCACCGTGCCCGCGCGGTCCAC 673

Qy     211 TACCCCATTGGGTGGGCTGTGCTCCGGCGCGCCTGGACTTACCGGATCCAGGAGGTGAGC 270
           | | ||  |   |||  |  || |     || |  |  |   |  ||  |||  ||
Db     672 GATCACACCGACCGGGTTCGGCACGACGACGGCGAGGATGCGCACGTCGCCGAGACGACG 613

Qy     271 GGGAGCTGCAATCTGCCTGCTG 292
           |   || ||   ||||| | ||
Db     612 GAACGCCGCGCCCTGCCCGGTG 591
```

RESULT 13
V50431/c
ID V50431 standard; DNA; 7193 BP.
AC V50431;
DT 26-OCT-1998 (first entry)
DE Streptomyces clavuligerus DNA sequence of ORFs flanking cas1.
KW Streptomyces clavuligerus; bacterial gene; clavulanic acid; 5S clavum;
KW 5R clavum; cas1; ORF; open reading frame; biosynthesis; ss.
OS Streptomyces clavuligerus.
PN WO9833896-A2.
PD 06-AUG-1998.
PF 02-FEB-1998; E00644.
PR 04-FEB-1997; GB-002218.
PA (SMIK) SMITHKLINE BEECHAM PLC.
PA (UYAL-) UNIV ALBERTA.
PI Anders C, Barton B, Griffin JP, Jensen S, Mosher RH,
PI Paradkar AS;
DR WPI; 98-437451/37.
PT DNA comprising defective 5S clavin biosynthesis gene(s) from
PT Streptomyces clavam - useful for producing clavulanic acid without
PT production of 5S clavam or clavam-2-carboxylates
PS Claim 2; Page 15-24; 29pp; English.
CC The present sequence represents DNA comprising genes specific for 5S
CC clavam biosynthesis from Streptomyces clavuligerus (SC) and which are
CC not essential for 5R clavam biosynthesis. The present invention also
CC describes: (1) a process for improving 5R clavam production in a
CC suitable microorganism comprising manipulation of DNA as above and its
CC inclusion in the microorganism; (2) a process for improving 5R clavam
CC production in SC comprising disrupting or otherwise making defective DNA
CC regions flanking cas 1; (3) a process for the identification of a
CC microorganism suitable for high 5R clavam production comprising a

```
CC   preliminary screening for microorganisms with low or no 5S clavam
CC   production; (4) a microorganism which is capable of 5R clavam production
CC   and low or no 5S clavam production obtainable by a process as in (3);
CC   (5) clavulanic acid (CA) obtainable by the fermentation of a
CC   microorganism as in (4), and (6) CA which is free of any 5S clavam; (7)
CC   CA which is free of any clavam-2-carboxylate. The methods and products
CC   can be used to produce organisms capable of producing increased amounts
CC   of clavams suitably e.g. CA, for use as antibiotics. The methods can
CC   also be used for the production of CA without the production of 5S
CC   clavam or clavam-2-carboxylate.
SQ   Sequence   7193 BP;    1055 A;    2745 C;    2498 G;    895 T;


  Query Match             3.6%;  Score 35.6;  DB 1;  Length 7193;
  Best Local Similarity   52.7%;  Pred. No. 1.9;
  Matches   77;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy      97 GCCTGCAGCATGAACCTGTGGCTCCTGGCCTGCCTGGTGGCCGGCTTCCTGGGAGCCTGG 156
           || | |||   |  || | || ||| || | ||| || |||  ||   |   |  | ||
Db    2735 GCGTACAGGTCGGCCCGGGGGGTCCGGGGCAGCCGGGCGGCGAGCGCGCGCAGCTCGGGG 2676

Qy     157 GCCCCCGCTGTCCACGCCCAAGGTGTCTTTGAGGACTGCTGCCTGGCCTACCACTACCCC 216
             |   ||| || ||  ||   |  ||| | | ||  || | |||  |||         |
Db    2675 CCGTACGCGGTGCAGTCCACTGAGGTCATGGTGGTGTGGTTCCTTTCCTTGGTTCCTCGT 2616

Qy     217 ATTGGGTGGGCTGTGCTCCGGCGCGC 242
             ||  | ||||||||  | || ||
Db    2615 GCCGGTGGTGCTGTGCTGTGCCGTGC 2590


RESULT  14
V26609
ID   V26609 standard; cDNA; 8169 BP.
AC   V26609;
DT   18-AUG-1998  (first entry)
DE   Actinomadura hibisca polyketide synthase gene.
KW   ds; multienzyme; infection; fungi; yeast; gram-positive bacteria; virus;
KW   dihydrobenzo(a)naphthacenequinone aglycon; antibiotic; pradimicin.
OS   Actinomadura hibisca.
FH   Key             Location/Qualifiers
FT   CDS             72. .1349
FT                   /*tag= a
FT                   /product= "Polyketide synthase unit"
FT   CDS             1346. .2569
FT                   /*tag= b
FT                   /product= "Polyketide synthase unit"
FT   CDS             2594. .2857
FT                   /*tag= c
FT                   /product= "Polyketide synthase unit"
FT   CDS             2854. .3315
FT                   /*tag= d
FT                   /product= "Polyketide synthase unit"
FT   CDS             3312. .3773
FT                   /*tag= e
FT                   /product= "Polyketide synthase unit"
FT   CDS             3794. .4819
FT                   /*tag= f
FT                   /product= "Polyketide synthase unit"
FT   CDS             4857. .5597
FT                   /*tag= g
FT                   /product= "Polyketide synthase unit"
FT   CDS             5594. .5935
FT                   /*tag= h
FT                   /product= "Polyketide synthase unit"
FT   CDS             5932. .6243
FT                   /*tag= i
FT                   /product= "Polyketide synthase unit"
FT   CDS             complement (7534. .6303)
FT                   /*tag= j
FT                   /product= "Polyketide synthase unit"
FT   CDS             7668. .8012
FT                   /*tag= k
FT                   /product= "Polyketide synthase unit"
PN   WO9811230-A1.
PD   19-MAR-1998.
PF   13-SEP-1996; U14791.
PR   13-SEP-1996; WO-U14791.
PA   (BRIM ) BRISTOL-MYERS SQUIBB CO.
PI   Dairi T, Oki T;
DR   WPI; 98-207391/18.
DR   P-PSDB; W54380, W54381, W54382, W54383, W54384, W54385, W54386, W54387;
DR   P-PSDB; W54388, W54389, W54390.
PT   Actinomadura polyketide synthase genes - useful for preparation of
PT   pradimicin
PS   Claim 4; Page 29-34; 71pp; English.
CC   The Actinomadura hibisca polyketide synthase gene encodes 11 separate
CC   proteins which are believed to form a multienzyme complex.  The enzyme is
CC   used for the biosynthesis of a dihydrobenzo(a)naphthacenequinone aglycon
CC   preferably a pradimicin which is an antibiotic useful against systemic
CC   fungal infections caused by Candida albicans, Aspergillus fumigatus and
CC   Cryptococcus neoformans.  It is also active against a wide variety of
CC   fungi and yeasts, some Gram-positive bacteria and viruses.
SQ   Sequence   8169 BP;    1119 A;    2971 C;    3015 G;    1064 T;


  Query Match             3.5%;  Score 35.2;  DB 1;  Length 8169;
  Best Local Similarity   49.5%;  Pred. No. 2.6;
  Matches   91;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy     102 CAGCATGAACCTGTGGCTCCTGGCCTGCCTGGTGGCCGGCTTCCTGGGAGCCTGGGCCCC 161
           | ||||| |||||  | | | ||||   |||  |       |    ||| | |  || |
Db    5213 CGGCATGTACCTGGTGGTGCGGGCCGCGCTGCCGCTGCTGTCGGAGGGCGGCGCGGTCGT 5272

Qy     162 CGCTGTCCACGCCCAAGGTGTCTTTGAGGACTGCTGCCTGGCCTACCACTACCCCATTGG 221
           ||  |||  | || | |  |   | |  | |  ||  | |  |  ||||||| ||   |
Db    5273 CGGCGTCGGCTCCAAGGTCGCGCTCGTCGGCATCTCGCAGCGCACCCACTACACCGCCGC 5332

Qy     222 GTGGGCTGTGCTCCGGCGCGCCTGGACTTACCGGATCCAGGAGGTGAGCGGGAGCTGCAA 281
             ||| | ||||    |   |  |   |  |  | | ||||| || |   |  | |||
Db    5333 CAAGGCCGGGCTCATCGGGTTCGTGCGCTCGCTCAGCAAGGAGCTGGGGCCGCTCGGCAT 5392

Qy     282 TCTG 285
           | |
Db    5393 CCGG 5396


RESULT  15
T78508
ID   T78508 standard; DNA; 44377 BP.
AC   T78508;
DT   26-FEB-1998  (first entry)
DE   Platenolide synthase gene cluster.
KW   Platenolide synthase gene cluster; platenolide production; srmG gene;
KW   multi-functional protein; macrolide antibiotic; spiramycin; ss.
OS   Streptomyces ambofaciens.
FH   Key             Location/Qualifiers
FT   CDS             350. .14002
FT                   /*tag= a
FT                   /transl_except= (pos:350. .352, aa:Met)
FT                   /note= "ORF1 encodes protein shown in W23716"
FT   CDS             14046. .20036
FT                   /*tag= b
FT                   /note= "ORF2 encodes protein shown in W23717"
FT   CDS             20110. .31284
FT                   /*tag= c
FT                   /transl_except= (pos:20111. .20113, aa:Met)
FT                   /note= "ORF3 encodes protein shown in W23718"
FT   CDS             31329. .36071
FT                   /*tag= d
FT                   /note= "ORF4 encodes protein shown in W23719"
FT   CDS             36155. .41830
FT                   /*tag= e
FT                   /note= "ORF5 encodes protein shown in W23720"
PN   EP-791656-A2.
PD   27-AUG-1997.
PF   19-FEB-1997; 301066.
```

PR 22-FEB-1996; US-012050.
PA (ELIL) LILLY & CO ELI.
PI Burgett SG, Kuhstoss SA, Rao RN, Richardson MA;
PI Rosteck PR;
DR WPI; 97-418047/39.
DR P-PSDB; W23716-W23720.
PT DNA encoding Streptomyces ambofaciens platenolide synthase domain -
PT for production of spiramycin-related polyketide antibiotics
PS Claim 9; Pages 8-33; 81pp; English.
CC This sequence represents the platenolide synthase gene cluster of the
CC invention. This sequence is referred to as the srmG gene, and was
CC isolated from Streptomyces ambofaciens. This sequence encodes the
CC multi-functional proteins which direct the synthesis of the polyketide
CC platenolide. Platenolide is the basic building block of the macrolide
CC antibiotic spiramycin. The DNA can be used to produce compounds
CC exhibiting antibiotic activity based on the platenolide structure,
CC including specifically the macrolide antibiotic spiramycin and spiramycin
CC analogues and derivatives. Modifications of the platenolide synthase DNA
CC sequence can be made so as to change the number and type of carboxylic
CC acids incorporated into the growing polyketide chain and to change the
CC kind of post-condensation processing that is conducted.
SQ Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T;

Query Match 3.5%; Score 34.6; DB 1; Length 44377;
Best Local Similarity 46.5%; Pred. No. 7.4;
Matches 112; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

```
Qy      1 GACAGCTTGGCCTACAGCCCGGCGGGCATCAGCTCCCTTGACCCAGTGGATATCGGTGGC 60
          ||  ||   | |  |  |   | || |  |  |  | |   |   || |   |||| | |
Db  29122 GAGGGCGCCGGCACCGTCGTCGAGGTCGGCCCCGGCGTCACCGGTGTCGCCGTCGGCGAC 29181

Qy     61 CCCGTTATTCGTCCAGGTGCCCAGGGAGGAGGACCCGCCTGCAGCATGAACCTGTGGCTC 120
          | |||  |  | |   | | |   ||    || ||    |||      |||   ||||
Db  29182 CGCGTGCTCGGCCTGTGGGACGGCGGCCTGGGCCCGCTGTGCGTGGCCGACCACCGGCTG 29241

Qy    121 CTGGCCTGCCTGGTGGCCGGCTTCCTGGGAGCCTGGGCCCCCGCTGTCCACGCCCAAGGT 180
          || |||  | |   || |||||        |||  |||| || | |||| |||
Db  29242 CTCGCCCCCGTCCCGGACGGCTGGTCCTACGCCCAGGCCGCCTCGGTCCCCGCGGTGTTC 29301

Qy    181 GTCTTTGAGGACTGCTGCCTGGCCTACCACTACCCCATTGGGTGGGCTGTGCTCCGGCGC 240
           ||   |   ||| | | |||| |  ||    |  | |  ||  ||  | ||    || |
Db  29302 CTCAGCGCCTACTACGGTCTGGTCACCCTGGCCGGCCTCAGGCCGGGGGAGCGGGTGCTC 29361

Qy    241 G 241
          |
Db  29362 G 29362
```

Search completed: July 24, 1999, 11:29:13
Job time: 14510 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: July 24, 1999, 11:28:19 ; Search time 16.5 Seconds
(without alignments)
183.867 Million cell updates/sec

Title: US-09-002-485-25
Perfect score: 710
Sequence: 1 MNLWLLACLVAGFLGAWAPA.....NPISSSKRNVSLLISANSGL 150

Scoring table: PAM150

Searched: 162890 seqs, 20225328 residues

Database : A_Geneseq_34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	708	99.7	150	1	W30565	Human chemokine ZS
2	698	98.3	150	1	W44397	Human thymus expre
3	689.5	97.1	149	1	W41938	Human chemokine be
4	312	43.9	144	1	W44396	Mouse thymus expre
5	113	15.9	133	1	W28511	Product of clone L
6	113	15.9	133	1	W50884	Amino acid sequenc
7	106	14.9	93	1	W27271	Novel beta-chemoki
8	106	14.9	94	1	W64475	Human eosinophil c
9	105	14.8	134	1	R81567	Chemokine beta-9.
10	102	14.4	134	1	W00668	Pancreas expressed
11	102	14.4	134	1	W50885	Amino acid sequenc
12	102	14.4	127	1	W60649	Human DVic-1 C-C c
13	102	14.4	134	1	W69163	CC-type chemokine
14	98	13.8	122	1	W25941	Mouse CCF18 chemok
15	98	13.8	122	1	W60018	Murine macrophage
16	96.5	13.6	79	1	W17664	Stem cell mobilisi
17	92.5	13.0	93	1	W14918	Mouse thymus and a
18	91.5	12.9	71	1	W64479	Human truncated eo
19	89.5	12.6	78	1	W64428	Human MPIF-1 splic
20	89.5	12.6	137	1	W57696	Human MPIF-1 splic
21	89.5	12.6	92	1	W64426	Human MPIF-1 splic
22	89.5	12.6	84	1	W64427	Human MPIF-1 splic
23	89	12.5	121	1	R76126	Macrophage inflamm
24	88	12.4	164	1	W60652	Human DVic-1 (alte
25	87.5	12.3	120	1	W07202	Human cytokine bet
26	87.5	12.3	137	1	W05186	Human eosinophil-e
27	87.5	12.3	82	1	W16315	Human chemokine be
28	87.5	12.3	76	1	W16317	Human chemokine be
29	87.5	12.3	75	1	W16318	Human chemokine be
30	87.5	12.3	77	1	W16316	Human chemokine be
31	87.5	12.3	82	1	W17663	Stem cell mobilisi
32	87.5	12.3	120	1	W57688	Human MPIF-1 prote
33	87.5	12.3	78	1	W57689	Human MPIF-1 prote
34	87.5	12.3	83	1	W57691	Human MPIF-1 prote
35	87.5	12.3	82	1	W57692	Human MPIF-1 prote
36	87.5	12.3	77	1	W57693	Human MPIF-1 prote
37	87.5	12.3	76	1	W57694	Human MPIF-1 prote
38	87.5	12.3	83	1	W57699	Human MPIF-1 prote
39	87.5	12.3	84	1	W57700	Human MPIF-1 prote
40	87.5	12.3	77	1	W57701	Human MPIF-1 prote
41	87.5	12.3	76	1	W57702	Human MPIF-1 prote
42	87.5	12.3	100	1	W57704	Human MPIF-1 prote
43	87.5	12.3	100	1	W57705	Human MPIF-1 prote
44	86.5	12.2	72	1	W57695	Human MPIF-1 prote
45	86.5	12.2	73	1	W57703	Human MPIF-1 prote

ALIGNMENTS

RESULT 1
W30565
ID W30565 standard; Protein; 150 AA.
AC W30565;
DT 02-FEB-1999 (first entry)
DE Human chemokine ZSIG-35.
KW ZSIG-35; beta-chemokine; human; ligand; lymphocyte migration;
KW inflammation; ischaemia; reperfusion injury.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1. .20
FT /label= Sig_peptide
FT Protein 21. .150
FT /label= Mat_protein
FT /note= "Claim 4"
PN WO9844117-A1.
PD 08-OCT-1998.
PF 27-MAR-1998; U06115.
PR 09-MAY-1997; US-046083.
PR 28-MAR-1997; US-042862.
PA (ZYMO) ZYMOGENETICS INC.
PI Sheppard PO;
DR WPI; 98-557114/47.
DR N-PSDB; V45444.
PT New human chemokine ZSIG-35 - used for, e.g. treating inflammatory
PT disease, lymphocyte migration and ischaemia/reperfusion injury
PS Claim 1; Page 80-81; 105pp; English.
CC This is the amino acid sequence of a novel human chemokine,
CC designated ZSIG-35, that has homology to members of the
CC beta-chemokine family, in particular to murine macrophage
CC inflammatory protein (35.8% identity). It includes a 4-Cys
CC motif (residues 30, 31, 58 and 75) that is characteristic of
CC beta-chemokines. The sequence was deduced from an isolated EST
CC clone (see V45444). The invention provides ZSIG-35 polypeptides
CC and polynucleotides, an expression vector, a cultured cell carrying
CC the vector, a method of producing the protein, a specific antibody,
CC a binding protein, and a pharmaceutical composition comprising
CC a ZSIG-35 polypeptide. ZSIG-35 may be used in therapeutics for
CC the regulation of acute and chronic inflammatory disease conditions,
CC lymphocyte migration and ischaemia/reperfusion injury.
SQ Sequence 150 AA;

Query Match 99.7%; Score 708; DB 1; Length 150;
Best Local Similarity 99.3%; Pred. No. 5.5e-81;
Matches 149; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MNLWLLACLVAGFLGAWAPAVHAQGVFEDCCLAYHYPIGWAVLRRAWTYRIQEVSGSCNL 60
           |||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db       1 MNLWLLACLVAGFLGAWAPAVHTQGVFEDCCLAYHYPIGWAVLRRAWTYRIQEVSGSCNL 60

Qy      61 PAAIFYLPKRHRKVCGNPKSREVQRAMKLLDARNKVFAKLRHNTQTFQAGPHAVKKLSSG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 PAAIFYLPKRHRKVCGNPKSREVQRAMKLLDARNKVFAKLRHNTQTFQAGPHAVKKLSSG 120

Qy     121 NSKLSSSKFSNPISSSKRNVSLLISANSGL 150
           |||||||||||||||||||||||||||||||
Db     121 NSKLSSSKFSNPISSSKRNVSLLISANSGL 150
```

RESULT 2
W44397
ID W44397 standard; Protein; 150 AA.
AC W44397;
DT 11-JUN-1998 (first entry)
DE Human thymus expressed chemokine.

```
KW   Human; thymus expressed chemokine; TECK; MIP-3alpha; MIP-3beta;
KW   receptor; dendritic cell; macrophage; inflammation; asthma.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Peptide         1. .23
FT                   /label= Signal
FT   Protein         24. .150
FT                   /label= TECK
PN   WO9801557-A2.
PD   15-JAN-1998.
PF   02-JUL-1997; U10819.
PR   04-JUN-1997; US-048593.
PR   05-JUL-1996; US-675814.
PR   11-OCT-1996; US-028329.
PA   (SCHE ) SCHERING CORP.
PI   Gish KC, Schall TJ, Vicari A, Wang W, Zlotnik A;
DR   WPI; 98-101054/09.
DR   N-PSDB; V15415.
PT   Novel chemokines, e.g. thymus expressed chemokine - used for
PT   treating inflammatory conditions including asthma.
PS   Claim 3; Page 88-89; 202pp; English.
CC   The present sequence represents a thymus expressed chemokine (TECK).
CC   Antibodies which bind to the protein can be used in detecting or
CC   diagnosing various immunological conditions related to expression
CC   of the protein. The nucleic acid can be used for screening and
CC   isolating DNA clones for the chemokines, especially from other
CC   species. The chemokine can be used in the treatment of conditions
CC   associated with abnormal physiology or development, including
CC   inflammatory conditions such as asthma.
SQ   Sequence   150 AA;

  Query Match             98.3%;  Score 698;  DB 1;  Length 150;
  Best Local Similarity   98.0%;  Pred. No. 9.8e-80;
  Matches  147;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 MNLWLLACLVAGFLGAWAPAVHAQGVFEDCCLAYHYPIGWAVLRRAWTYRIQEVSGSCNL 60
           |||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db       1 MNLWLLACLVAGFLGAWAPAVHTQGVFEDCCLAYHYPIGWAVLRRAWTYRIQEVSGSCNL 60

Qy      61 PAAIFYLPKRHRKVCGNPKSREVQRAMKLLDARNKVFAKLRHNTQTFQAGPHAVKKLSSG 120
           ||||||||||||||||||||||||||||||||||||||:|| ||||||||||||||||
Db      61 PAAIFYLPKRHRKVCGNPKSREVQRAMKLLDARNKVFAKLHHNMQTFQAGPHAVKKLSSG 120

Qy     121 NSKLSSSKFSNPISSSKRNVSLLISANSGL 150
           |||||||||||||||||||||||||||||
Db     121 NSKLSSSKFSNPISSSKRNVSLLISANSGL 150


RESULT   3
W41938
ID   W41938 standard; Protein; 149 AA.
AC   W41938;
DT   22-JUN-1998  (first entry)
DE   Human chemokine beta-15.
KW   Chemokine beta-15; CB15; human; thymus; thymocyte; inflammation;
KW   autoimmune disease; immunosuppression; tumour; cancers;
KW   hypoactivity; hyperactivity; atrophy; arthritis; leukaemia;
KW   lymphomas; sepsis; wound healing; myelosuppression; diagnosis;
KW   therapy.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Peptide         1. .20
FT                   /label= Sig_peptide
FT                   /note= "21"
FT   Protein         21. .149
FT                   /label= Mat_protein
PN   WO9748807-A1.
PD   24-DEC-1997.
PF   17-JUN-1996; U10561.
PR   17-JUN-1996; WO-U10561.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Kreider BL, Rosen CA, Wei Y;
DR   WPI; 98-063145/06.
DR   N-PSDB; V04481.
PT   New human chemokine beta-15 gene - used to develop products for the
PT   diagnosis and treatment of thymus-related disorders, e.g.
PT   auto-immunity, immunosuppression or inflammation
PS   Claim 1; Page 60; 79pp; English.
CC   This polypeptide comprises human chemokine beta-15 (CB15), a novel
CC   CC chemokine, expressed only in thymus tissue, that modulates early
CC   thymocyte proliferation and differentiation, mediates the
CC   differentiation of intrathymic T cell precursors into mature
CC   T-lymphocytes, and modulates colony formation of bone marrow
CC   progenitor cells.  The CB15 amino acid sequence was deduced from
CC   cDNA clone HTSEX82 (ATCC 97519) (see V41938).  It shows 34%
CC   identity and 53% similarity to mouse macrophage inflammatory
CC   protein-related protein 2 (MMRP-2).  Assay of CB15 expression can
CC   be used in methods for the diagnosis of disorders of the thymus
CC   such as tumours, cancers, hypoactivity, hyperactivity, atrophy,
CC   enlargement of the thymus, or disorders involving autoimmunity,
CC   arthritis, leukaemias, lymphomas, immunosuppression, sepsis, wound
CC   healing, acute and chronic inflammation, cell mediated immunity,
CC   humor immunity, or TH1/TH2 imbalance.  CB15 polypeptides can be
CC   expressed in host cells and used to raise CB15-specific antibodies
CC   and to treat individuals in need of increased levels of CB15 e.g.
CC   for myelosuppression.
SQ   Sequence   149 AA;

  Query Match             97.1%;  Score 689.5;  DB 1;  Length 149;
  Best Local Similarity   98.0%;  Pred. No. 1.1e-78;
  Matches  147;  Conservative    2;  Mismatches    0;  Indels    1;  Gaps    1;

Qy       1 MNLWLLACLVAGFLGAWAPAVHAQGVFEDCCLAYHYPIGWAVLRRAWTYRIQEVSGSCNL 60
           |||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db       1 MNLWLLACLVAGFLGAWAPAVHTQGVFEDCCLAYHYPIGWAVLRRAWTYRIQEVSGSCNL 60

Qy      61 PAAIFYLPKRHRKVCGNPKSREVQRAMKLLDARNKVFAKLRHNTQTFQAGPHAVKKLSSG 120
           ||||||||||||||||||||||||||||||||||||||:||||||| ||||||||||
Db      61 PAAIFYLPKRHRKVCGNPKSREVQRAMKLLDARNKVFAKLHHNTQTFQ-GPHAVKKLSSG 119

Qy     121 NSKLSSSKFSNPISSSKRNVSLLISANSGL 150
           |||||||||||||||||||||||||||||
Db     120 NSKLSSSKFSNPISSSKRNVSLLISANSGL 149


RESULT   4
W44396
ID   W44396 standard; Protein; 144 AA.
AC   W44396;
DT   11-JUN-1998  (first entry)
DE   Mouse thymus expressed chemokine.
KW   Mouse; thymus expressed chemokine; TECK; MIP-3alpha; MIP-3beta;
KW   receptor; dendritic cell; macrophage; inflammation; asthma.
OS   Mus sp.
FH   Key             Location/Qualifiers
FT   Peptide         1. .23
FT                   /label= Signal
FT                   /note= "Putative"
FT   Protein         24. .144
FT                   /label= TECK
PN   WO9801557-A2.
PD   15-JAN-1998.
PF   02-JUL-1997; U10819.
PR   04-JUN-1997; US-048593.
PR   05-JUL-1996; US-675814.
PR   11-OCT-1996; US-028329.
PA   (SCHE ) SCHERING CORP.
PI   Gish KC, Schall TJ, Vicari A, Wang W, Zlotnik A;
DR   WPI; 98-101054/09.
DR   N-PSDB; V15414.
PT   Novel chemokines, e.g. thymus expressed chemokine - used for
PT   treating inflammatory conditions including asthma.
PS   Claim 3; Page 86-87; 202pp; English.
CC   The present sequence represents a thymus expressed chemokine (TECK).
```

```
CC   Antibodies which bind to the protein can be used in detecting or
CC   diagnosing various immunological conditions related to expression
CC   of the protein. The nucleic acid can be used for screening and
CC   isolating DNA clones for the chemokines, especially from other
CC   species. The chemokine can be used in the treatment of conditions
CC   associated with abnormal physiology or development, including
CC   inflammatory conditions such as asthma.
SQ   Sequence   144 AA;


  Query Match             43.9%;  Score 312;  DB 1;  Length 144;
  Best Local Similarity   48.6%;  Pred. No. 1.5e-31;
  Matches   69;  Conservative   28;  Mismatches   33;  Indels   12;  Gaps    4;

Qy       1 MNLWLLACLVAGFLGAWAPAVHAQGVFEDCCLAYHYPIGWAVLRRAWTYRIQEVSGSCNL 60
           |:|||:|||| |:||| | ||||| ||||||:|:  | | |||:|: |: |||||||||
Db       1 MKLWLFACLVACFVGAWMPVVHAQGAFEDCCLGYQHRIKWNVLRHARNYHQQEVSGSCNL 60

Qy      61 PAAIFYLPKRHRKVCGNPKSREVQRAMKLLDARNKVFAKLRHNTQTFQAGPHAVKKLSSG 120
             |  ||:  |:: |||||   :| ||:::| ||:    :| |   :  : ::: :|
Db      61 RAVRFYF--RQKVVCGNPEDMNVKRAIRILTARK----RLVHWKSA--SDSQTERK---- 109

Qy     121 NSKLSSSKFSNPISSSKRNVSL 142
           :|:   ||  || |:| |: :|
Db     109 KSNHMKSKVENPNSTSVRSATL 130


RESULT   5
W28511
ID   W28511 standard; Protein; 133 AA.
AC   W28511;
DT   29-DEC-1997  (first entry)
DE   Product of clone L105.
KW   J5; J422; L105; H174-10; H174-43; B18; cytokine; PBMC;
KW   peripheral blood mononuclear cell; disintegrin; metallo-protein;
KW   Drosophila; leucine-rich repeat; monocyte; chemoattractant;
KW   IP-10; CRG-2; CTLA-8; herpesvirus; Saimiri.
OS   Mus musculus.
PN   WO9707198-A2.
PD   27-FEB-1997.
PF   08-AUG-1996; U12897.
PR   08-AUG-1996; WO-U12897.
PA   (GEMY ) GENETICS INST INC.
PI   Carlin M,  Jacobs K,  Kelleher K,  Mccoy JM;
DR   WPI; 97-165283/15.
DR   N-PSDB; T87429.
PT   Polynucleotide(s) encoding proteins for treating, preventing and
PT   ameliorating medical conditions - obtained from human activated
PT   peripheral blood mononuclear cell, and murine adult thymus libraries
PS   Claim 21; Page 44-45; 61pp; English.
CC   This sequence was isolated from a murine adult thymus library using
CC   a trap selecting for nucleotides encoding secreted proteins, and
CC   encodes a protein having homology to various monocyte and other
CC   chemoattractant proteins.
SQ   Sequence   133 AA;


  Query Match             15.9%;  Score 113;  DB 1;  Length 133;
  Best Local Similarity   37.2%;  Pred. No. 9.8e-07;
  Matches   35;  Conservative   21;  Mismatches   28;  Indels   10;  Gaps    6;

Qy       1 MNLWLLACLVAGFLGAWAPAVHAQGVFEDCCLAY-HYPIGWAVLRRAWTYRIQEVSGSCN 59
           | | |: || ::  :|: :  : |  :|||| | :  |  :::|    || || | :|
Db       5 MTLSLLS-LVLALCIPWTQGSDGGG--QDCCLKYSQKKIPYSIVR---GYRKQEPSLGCP 58

Qy      60 LPAAIFYLPKRHRK--VCGNPKSREVQRAMKLLD 91
           :| ||:: |::| |  :|:||    ||  |: ||
Db      59 IP-AILFSPRKHSKPELCANPEEGWVQNLMRRLD 91


RESULT   6
W50884
ID   W50884 standard; Protein; 133 AA.
AC   W50884;
DT   09-SEP-1998  (first entry)
DE   Amino acid sequence of mouse 6CKine protein.
KW   Mouse; 6CKine gene; m6CKine; chemokine; mpf4; mCTAP3;
KW   h6Ckine; Chr19kine; cancer; degenerative condition; antibody;
KW   immuno assay; forensic assay; in situ assay.
OS   Mus sp.
FH   Key             Location/Qualifiers
FT   Peptide         1. .23
FT                   /note= "signal peptide"
FT   Protein         24. .133
FT                   /note= "mature protein"
PN   WO9814581-A1.
PD   09-APR-1998.
PF   02-OCT-1997; U17122.
PR   28-AUG-1997; US-058007.
PR   02-OCT-1996; US-027242.
PR   09-OCT-1996; US-028042.
PA   (SCHE ) SCHERING CORP.
PI   Hedrick JA, Zlotnik A;
DR   WPI; 98-240086/21.
DR   N-PSDB; V07113.
PT   Mouse and human CC and CXC chemokine(s) - useful to modulate
PT   physiology or development of cells to treat, e.g. cancerous or
PT   degenerative conditions
PS   Claim 1; Pages 78-79; 88pp; English.
CC   This is the amino acid sequence of the mouse 6CKine (m6CKine) gene, a
CC   chemokine.  It is used in the method of the invention where mouse and
CC   human CC and CXC chemokines, designated mpf4, mCTAP3, m6Ckine, h6Ckine
CC   and Chr19kine are used to modulate the physiology or the development
CC   of cells to treat, cancerous or degenerative conditions.  The
CC   chemokines can also be used to generate antibodies, useful in
CC   immunoassays to measure chemokines, while the nucleic acid sequences
CC   may be used as components in forensic assays or in in situ assays to
CC   detect chromosomal abnormalities.
SQ   Sequence   133 AA;


  Query Match             15.9%;  Score 113;  DB 1;  Length 133;
  Best Local Similarity   37.2%;  Pred. No. 9.8e-07;
  Matches   35;  Conservative   21;  Mismatches   28;  Indels   10;  Gaps    6;

Qy       1 MNLWLLACLVAGFLGAWAPAVHAQGVFEDCCLAY-HYPIGWAVLRRAWTYRIQEVSGSCN 59
           | | ||: || ::  :|: :  : |  :|||| | :  |  :::|    || || | :|
Db       5 MTLSLLS-LVLALCIPWTQGSDGGG--QDCCLKYSQKKIPYSIVR---GYRKQEPSLGCP 58

Qy      60 LPAAIFYLPKRHRK--VCGNPKSREVQRAMKLLD 91
           :| ||:: |::| |  :|:||    ||  |: ||
Db      59 IP-AILFSPRKHSKPELCANPEEGWVQNLMRRLD 91


RESULT   7
W27271
ID   W27271 standard; Peptide; 93 AA.
AC   W27271;
DT   27-APR-1998  (first entry)
DE   Novel beta-chemokine designated PTEC.
KW   Beta-chemokine; PTEC; homologue; human RANTES; C-C chemokine;
KW   diagnosis; treatment; activated T cell;
KW   immediate type hypersensitivity; leukocyte proliferation.
OS   Homo sapiens.
PN   WO9739126-A1.
PD   23-OCT-1997.
PF   14-APR-1997; U06249.
PR   17-APR-1996; US-633682.
PA   (INCY-) INCYTE PHARM INC.
PI   Bandman O, Hawkins PR, Murry LE;
DR   WPI; 97-526462/48.
DR   N-PSDB; T91275.
PT   PTEC, a novel human beta-chemokine - useful for diagnosing and
PT   treating e.g. viral, bacterial, fungal infections, auto-immune
PT   diseases, etc
PS   Claim 1; Page 40; 59pp; English.
```

```
CC   The present sequence represents a novel beta-chemokine polypeptide
CC   designated PTEC. The protein has Cys residues of a beta-chemokine (C32,
CC   C33, C56 and C72). The cDNA sequence encoding PTEC was isolated from a
CC   cDNA library generated from non-tumourous human prostate tissue. The
CC   PTEC protein is a homologue of human RANTES, which is a C-C chemokine
CC   with structural similarity to interleukin-8 and human MIP-1beta. The
CC   polynucleotide sequence encoding PTEC can be used to diagnose or treat
CC   a condition associated with the presence of activated T cells including
CC   viral, bacterial, fungal or helminthic infections, allergic or asthmatic
CC   responses, mechanical injury associated with trauma, arteriosclerosis,
CC   atherogenesis or collagen vascular diseases, autoimmune diseases such as
CC   rheumatoid arthritis, myasthenia gravis, systemic lupus erythematosus or
CC   haemolytic anaemia, leukaemia, lymphomas or carcinomas, and diseases of
CC   immediate type hypersensitivity which involve activation or excessive
CC   proliferation of leukocytes, particularly monocytes, macrophages,
CC   eosinophils, basophil, mast and T cells.
SQ   Sequence   93 AA;


  Query Match            14.9%;  Score 106;  DB 1;  Length 93;
  Best Local Similarity  32.3%;  Pred. No. 4.7e-06;
  Matches   30;  Conservative   20;  Mismatches   33;  Indels   10;  Gaps    4;

Qy       5 LLACLVAGFLGAWAPAVHAQGVFEDCCLAY-HYPIGWAVLRRAWTYRIQEVSGSCNLPAA 63
           ||| |::  ||:   | ::  :   ||: | | |: |     :|    :  | ||:  |
Db      10 LLASLLSLHLGT---ATRGSDISKTCCFQYSHKPLPW-----TWVRSYEFTSNSCSQRAV 61

Qy      64 IFYLPKRHRKVCGNPKSREVQRAMKLLDARNKV 96
           ||   || :||| :|: : ||: : || : : :
Db      62 IF-TTKRGKKVCTHPRKKWVQKYISLLKTPKQL 93


RESULT   8
W64475
ID   W64475 standard; Protein; 94 AA.
AC   W64475;
DT   18-NOV-1998  (first entry)
DE   Human eosinophil cell activator HVC002 protein.
KW   Eosinophil cell activator; treatment; diagnosis; malignant tumour;
KW   parasitic infection; allergic inflammation; eosinophilic pneumonia;
KW   rapid onset eosinophilia; autoimmune disease; gene therapy.
OS   Homo sapiens.
PN   WO9824817-A1.
PD   11-JUN-1998.
PF   05-DEC-1997; J04470.
PR   05-DEC-1996; JP-325762.
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PI   Koike M, Kuga T, Nakagawa S, Nishi T, Saito A,
PI   Shinkai A, Yoshisue H;
DR   WPI; 98-333261/29.
DR   N-PSDB; V49501, V49508.
PT   DNA and encoded protein which activates eosinophil cells - for
PT   treatment of cancer, parasite infection, autoimmune disease and
PT   allergic inflammation
PS   Claim 2; Page 63; 92pp; Japanese.
CC   This sequence represents a human eosinophil cell activator. This protein
CC   and antibodies generated from the protein can be used for treatment and
CC   diagnosis of malignant tumours, parasitic infections, allergic
CC   inflammation, eosinophilic pneumonia, rapid onset eosinophilia, and
CC   autoimmune diseases. DNA can be used for diagnosis, and the antisense
CC   DNA in gene therapy of these disorders. The protein can be used for
CC   screening of potential agonists or antagonists of its activity.
SQ   Sequence   94 AA;


  Query Match            14.9%;  Score 106;  DB 1;  Length 94;
  Best Local Similarity  32.3%;  Pred. No. 4.8e-06;
  Matches   30;  Conservative   20;  Mismatches   33;  Indels   10;  Gaps    4;

Qy       5 LLACLVAGFLGAWAPAVHAQGVFEDCCLAY-HYPIGWAVLRRAWTYRIQEVSGSCNLPAA 63
           ||| |::  ||:   | ::  :   ||: | | |: |     :|    :  | ||:  |
Db      11 LLASLLSLHLGT---ATRGSDISKTCCFQYSHKPLPW-----TWVRSYEFTSNSCSQRAV 62

Qy      64 IFYLPKRHRKVCGNPKSREVQRAMKLLDARNKV 96
           ||   || :||| :|: : ||: : || : : :
Db      63 IF-TTKRGKKVCTHPRKKWVQKYISLLKTPKQL 94


RESULT   9
R81567
ID   R81567 standard; Protein; 134 AA.
AC   R81567;
DT   01-OCT-1996  (first entry)
DE   Chemokine beta-9.
KW   Human; chemokine beta-9; Ck beta-9; breast lymph node; chemokine;
KW   eotaxin; inhibition; bone marrow stem cell colony formation;
KW   cancer chemotherapy; leukaemia; epidermal keratinocyte proliferation;
KW   psoriasis; solid tumour; autoimmune disease; IL-2 biosynthesis.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   peptide         1. .23
FT                   /note= "Leader sequence"
FT   protein         24. .134
FT                   /note= "Mature Ck beta-9"
PN   WO9606169-A1.
PD   29-FEB-1996.
PF   06-JUN-1995; U06260.
PR   23-AUG-1994; US-294251.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Adams MD,  Li H;
DR   WPI; 96-151372/15.
DR   N-PSDB; T18015.
PT   Human chemokine beta-9 polynucleotide(s), polypeptide(s) and
PT   antagonists - useful in treatment of e.g. leukaemia, tumours,
PT   chronic infections and auto-immune disorders and diagnosis of Ck
PT   beta-9 mutation(s)
PS   Claim 12; Page 41; 55pp; English.
CC   This sequence represents human chemokine beta-9 (Ck beta-9).  The
CC   polynucleotide encoding CK beta-9 was discovered in a cDNA library
CC   derived from a human breast lymph node.  Ck beta-9 is structurally
CC   related to the chemokine family.  It exhibits the highest degree of
CC   homology to eotaxin with 32% identity and 69% similarity over a
CC   stretch of 75 amino acids.  The four spatially conserved cysteine
CC   residues found in chemokines are also found in Ck beta-9.  Ck beta-9
CC   polypeptides may be used to inhibit bone marrow stem cell colony
CC   formation as adjunct protective treatment during cancer chemotherapy and
CC   for leukaemia.  They can also be used to inhibit epidermal keratinocyte
CC   proliferation for treatment of psoriasis.  They may be used to treat
CC   solid tumours by stimulating the invasion and activation of host defence
CC   cells, e.g. cytotoxic T cells and macrophages.  They act to enhance host
CC   defences against resistant chronic infections, and to treat autoimmune
CC   disease and lymphocytic leukaemia by inhibiting T cell proliferation by
CC   the inhibition of IL-2 biosynthesis.
SQ   Sequence   134 AA;


  Query Match            14.8%;  Score 105;  DB 1;  Length 134;
  Best Local Similarity  36.2%;  Pred. No. 9.9e-06;
  Matches   34;  Conservative   24;  Mismatches   26;  Indels   10;  Gaps    6;

Qy       1 MNLWLLACLVAGFLGAWAPAVHAQGVFEDCCLAY-HYPIGWAVLRRAWTYRIQEVSGSCN 59
           : | ||  || :|  ::: :  ::|  :|||| | :  |   |:|   :|| || | :|:
Db       5 LALSLLI-LVLAFGIPRTQG--SDGGAQDCCLKYSQRKIPAKVVR---SYRKQEPSLGCS 58

Qy      60 LPAAIFYLPKR--HRKVCGNPKSREVQRAMKLLD 91
           :| ||::||::  :  :|::||   ||: |  ||
Db      59 IP-AILFLPRKRSQAELCADPKELYVQQLMQHLD 91


RESULT  10
W00668
ID   W00668 standard; Protein; 134 AA.
AC   W00668;
DT   02-MAY-1997  (first entry)
DE   Pancreas expressed chemokine-2.
KW   Pancreas-derived chemokine; PANEC-1; PANEC-2; diagnosis;
```

KW inflammation; disease; cancer.
OS Homo sapiens.
PN WO9625497-A1.
PD 22-AUG-1996.
PF 16-FEB-1996; U02225.
PR 17-FEB-1995; US-390740.
PA (INCY-) INCYTE PHARM INC.
PI Bandman O, Coleman R, Wilde CG;
DR WPI; 96-393398/39.
DR N-PSDB; T33528.
PT Nucleotide and protein sequences for human PANEC-1 and PANEC-2 -
PT useful in diagnosis and therapy of pancreatic diseases
PS Claim 20; Page 30; 43pp; English.
CC The sequences given in W00667-68 represent pancreas-derived chemokines,
CC PANEC-1 and PANEC-2. These chemokines are highly expressed and
CC specifically expressed in the pancreas and may therefore be used in
CC diagnostic assays based on chemokine production in cases of
CC inflammation or disease affecting the pancreas. These assays allow
CC the early and accurate diagnosis of pancreatic disorders, and can
CC differentiate between invasive diseases and genetic syndromes.
SQ Sequence 134 AA;

Query Match 14.4%; Score 102; DB 1; Length 134;
Best Local Similarity 36.2%; Pred. No. 2.3e-05;
Matches 34; Conservative 24; Mismatches 26; Indels 10; Gaps 6;

Qy 1 MNLWLLACLVAGFLGAWAPAVHAQGVFEDCCLAY-HYPIGWAVLRRAWTYRIQEVSGSCN 59
: | || || :| ::: : ::| :|||| | : | |:| :|| || | :|:
Db 5 LALSLLI-LVLAFGIPRTQG--SDGGAQDCCLKYSQRKIPAKVVR---SYRKQEPSLGCS 58

Qy 60 LPAAIFYLPKR--HRKVCGNPKSREVQRAMKLLD 91
:| ||::||:: : :|::|| ||: | ||
Db 59 IP-AILFLPRKRSQAELCADPKELWVQQLMQHLD 91

RESULT 11
W50885
ID W50885 standard; Protein; 134 AA.
AC W50885;
DT 09-SEP-1998 (first entry)
DE Amino acid sequence of human 6CKine protein.
KW Human; 6CKine gene; h6CKine; chemokine; mpf4; mCTAP3;
KW m6Ckine; Chr19kine; cancer; degenerative condition; antibody;
KW immuno assay; forensic assay; in situ assay.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1. .23
FT /note= "signal peptide"
FT Protein 24. .134
FT /note= "mature protein"
PN WO9814581-A1.
PD 09-APR-1998.
PF 02-OCT-1997; U17122.
PR 28-AUG-1997; US-058007.
PR 02-OCT-1996; US-027242.
PR 09-OCT-1996; US-028042.
PA (SCHE) SCHERING CORP.
PI Hedrick JA, Zlotnik A;
DR WPI; 98-240086/21.
DR N-PSDB; V07114.
PT Mouse and human CC and CXC chemokine(s) - useful to modulate
PT physiology or development of cells to treat, e.g. cancerous or
PT degenerative conditions
PS Claim 1; Page 81; 88pp; English.
CC This is the amino acid sequence of the human 6CKine (h6CKine) protein,
CC a chemokine. It is used in the method of the invention where mouse
CC and human CC and CXC chemokines, designated mpf4, mCTAP3, m6Ckine,
CC h6Ckine and Chr19kine are used to modulate the physiology or the
CC development of cells to treat, cancerous or degenerative conditions.
CC The chemokines can also be used to generate antibodies, useful in
CC immunoassays to measure chemokines, while the nucleic acid sequences
CC may be used as components in forensic assays or in in situ assays to
CC detect chromosomal abnormalities.
SQ Sequence 134 AA;

Query Match 14.4%; Score 102; DB 1; Length 134;
Best Local Similarity 36.2%; Pred. No. 2.3e-05;
Matches 34; Conservative 24; Mismatches 26; Indels 10; Gaps 6;

Qy 1 MNLWLLACLVAGFLGAWAPAVHAQGVFEDCCLAY-HYPIGWAVLRRAWTYRIQEVSGSCN 59
: | || || :| ::: : ::| :|||| | : | |:| :|| || | :|:
Db 5 LALSLLI-LVLAFGIPRTQG--SDGGAQDCCLKYSQRKIPAKVVR---SYRKQEPSLGCS 58

Qy 60 LPAAIFYLPKR--HRKVCGNPKSREVQRAMKLLD 91
:| ||::||:: : :|::|| ||: | ||
Db 59 IP-AILFLPRKRSQAELCADPKELWVQQLMQHLD 91

RESULT 12
W60649
ID W60649 standard; Protein; 127 AA.
AC W60649;
DT 12-OCT-1998 (first entry)
DE Human DVic-1 C-C chemokine.
KW DVic-1; DNAX Vic-1; C-C chemokine; cytokine; human; immune system;
KW cancer; cell proliferation; therapy; diagnosis.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1. .22
FT /label= Sig_peptide
FT Protein 23. .127
FT /label= Mat_protein
FT /note= "Claim 1"
PN WO9823750-A2.
PD 04-JUN-1998.
PF 26-NOV-1997; U21092.
PR 05-DEC-1996; US-761071.
PR 27-NOV-1996; US-031805.
PA (SCHE) SCHERING CORP.
PI Hedrick JA, Morales J, Vicari A, Zlotnik A;
DR WPI; 98-322730/28.
DR N-PSDB; V38291.
PT DVic-1 and DGWCC chemokines - useful for developing products for
PT treating abnormal physiology or development, e.g. cancerous or
PT degenerative conditions
PS Claim 1; Page 59-60; 71pp; English.
CC This polypeptide comprises human DNAX Vic-1 (DVic-1), a novel C-C
CC chemokine, the mature portion of which is claimed. The amino acid
CC sequence was deduced from a cDNA clone (see V38291). An alternative
CC longer transcript (see W60652) for human DVic-1 is also disclosed.
CC Also claimed is novel human DNAX Groin Wound expressed CC chemokine
CC (DGWCC) (see W60649) mature protein, as well as expression vectors
CC and host cells. DVic-1 and DGWCC play a role in the regulation or
CC development of neuronal or haematopoietic cells, e.g. lymphoid
CC cells, which affect immunological responses. They can be used in
CC the treatment of conditions associated with abnormal physiology or
CC development, including abnormal proliferation, e.g. cancerous
CC conditions or degenerative conditions. Abnormal proliferation,
CC regeneration, degeneration, and atrophy may be modulated by
CC appropriate therapeutic treatment using products of the invention.
CC The products can also be used for detection, diagnosis and drug
CC screening.
SQ Sequence 127 AA;

Query Match 14.4%; Score 102; DB 1; Length 127;
Best Local Similarity 30.3%; Pred. No. 2.2e-05;
Matches 20; Conservative 19; Mismatches 25; Indels 2; Gaps 1;

Qy 30 CCLAYHYPIGWAVLRRAWTYRIQEVSGSCNLPAAIFYLPKRHRKVCGNPKSREVQRAMKL 89
|| :: :| | ||| | |:|:| |: : ::|::| :| :: | : ||:
Db 30 CCTEVSHHISRRLLERVNMCRIQRADGDCDLAAVILHV--KRRRICVSPHNHTVKQWMKV 87

Qy 90 LDARNK 95

```
          |:::
Db     88 QAAKKN 93
```

```
RESULT  13
W69163
ID   W69163 standard; Protein; 134 AA.
AC   W69163;
DT   15-OCT-1998  (first entry)
DE   CC-type chemokine protein designated SLC.
KW   Human; CC-type chemokine protein; SLC; secondary lymphoid chemokine;
KW   CCR 7 receptor; secondary lymphoid tissue; diagnosis; treatment;
KW   disease; inflammatory; immune-related; cancer; HIV.
OS   Homo sapiens.
PN   WO9831809-A1.
PD   23-JUL-1998.
PF   19-JAN-1998; J00154.
PR   20-JAN-1997; JP-007602.
PA   (SHIO ) SHIONOGI & CO LTD.
PI   Imai T, Nagira M, Yoshie O;
DR   WPI; 98-414107/35.
DR   N-PSDB; V41199.
PT   Antibodies to secondary lymphoid chemokine bind specifically to CCR
PT   7 receptor - useful in diagnosis and treatment of inflammatory,
PT   immune-related, infectious and neoplastic disorders
PS   Claim 2; Fig 1; 41pp; Japanese.
CC   The present sequence represents a human CC-type chemokine protein
CC   designated SLC (secondary lymphoid chemokine). This protein binds to the
CC   CCR 7 receptor and is expressed in secondary lymphoid tissue. Potential
CC   agonists, antagonists and inverse agonists to the receptor (CCR 7) which
CC   binds to SLC, can be screened using SLC or its derivatives or fragments,
CC   for possible therapeutic use. The products can also be used for diagnosis
CC   and treatment of diseases involving the interaction of SLC with the CCR 7
CC   receptor. These diseases comprise inflammatory, immune-related and
CC   infectious (including HIV) diseases and cancer.
SQ   Sequence   134 AA;
```

```
  Query Match             14.4%;  Score 102;  DB 1;  Length 134;
  Best Local Similarity   36.2%;  Pred. No. 2.3e-05;
  Matches   34;  Conservative   24;  Mismatches   26;  Indels   10;  Gaps    6;

Qy       1 MNLWLLACLVAGFLGAWAPAVHAQGVFEDCCLAY-HYPIGWAVLRRAWTYRIQEVSGSCN 59
           : | ||  || :|  ::: :  ::|  :|||| | :  |   |:|   :|| || | :|:
Db       5 LALSLLI-LVLAFGIPRTQG--SDGGAQDCCLKYSQRKIPAKVVR---SYRKQEPSLGCS 58

Qy      60 LPAAIFYLPKR--HRKVCGNPKSREVQRAMKLLD 91
           :| ||::||::  :  :|::||   ||: |  ||
Db      59 IP-AILFLPRKRSQAELCADPKELWVQQLMQHLD 91
```

```
RESULT  14
W25941
ID   W25941 standard; Protein; 122 AA.
AC   W25941;
DT   13-MAR-1998  (first entry)
DE   Mouse CCF18 chemokine.
KW   CCF18 chemokine; mouse; primer; PCR; amplification; antagonist;
KW   abnormal physiology; development; anti-viral.
OS   Mus sp.
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "signal peptide"
FT   Protein         22. .122
FT                   /note= "mature protein"
FT   Modified_site   98. .100
FT                   /note= "Asn is N-glycosylated"
PN   WO9721812-A2.
PD   19-JUN-1997.
PF   05-DEC-1996; U19139.
PR   08-DEC-1995; US-567882.
PA   (SCHE ) SCHERING CORP.
PI   Dairaghi DJ,  Hara T,  Miyajima A,  Schall TJ,  Wang W;
PI   Yoshimura A;
DR   WPI; 97-332784/30.
DR   N-PSDB; T79093.
PT   New isolated chemokine CCF8 and chemokine receptor CCKR3 - used to
PT   develop products useful for the diagnosis and treatment of
PT   conditions associated with abnormal physiology or development
PS   Claim 2; Page 57-58; 73pp; English.
CC   This is the amino acid sequence of a novel CCF18 chemokine isolated from
CC   mice.  The encoding sequence was isolated from a cDNA library generated
CC   from mRNA from epidermal growth factor (EGF)-stimulated BF-EGFR/EPORH
CC   cells, Ba/F3 (mouse pre-B-cell line) transfectants expressing a chimeric
CC   receptor comprising the EGF extracellular domain fused to the cytoplasmic
CC   domain of the erythropoietin receptor.  The cells are stimulated and
CC   proliferate in response to either interleukin-3 or EGF.  The chemokine
CC   comprises a mature protein of 102 amino acids with a molecular weight of
CC   about 11549 daltons.  The sequence contains 4 conserved Cys residues at
CC   positions 57, 58, 80 and 96, placing the chemokine in the CC group.
CC   The protein can be used to screen for (ant)agonists that bind to the
CC   chemokine.  These (ant)agonists are useful in the treatment of conditions
CC   associated with abnormal physiology or development.  The chemokine may
CC   also have antiviral activity.
SQ   Sequence   122 AA;
```

```
  Query Match             13.8%;  Score 98;  DB 1;  Length 122;
  Best Local Similarity   30.9%;  Pred. No. 6.6e-05;
  Matches   34;  Conservative   11;  Mismatches   19;  Indels   46;  Gaps    6;

Qy      14 LGAWAPAVHA------------QGV--------FE---DCCLAYHYPIGWAVLRRAWTY 49
           || ||   ||            ||:        |:   ||||:|:
Db      17 LGIWAQITHATETKEVQSSLKAQQGLEIEMFHMGFQDSSDCCLSYN------------S 63

Qy      50 RIQ--------EVSGSCNLPAAIFYLPKRHRKVCGNPKSREVQRAMKLLD 91
           |||          ||:|  |: || ::||   ||:||  | ||| :  |:
Db      64 RIQCSRFIGYFPTSGGCTRPGIIF-ISKRGFQVCANPSDRRVQRCIERLE 112
```

```
RESULT  15
W60018
ID   W60018 standard; Protein; 122 AA.
AC   W60018;
DT   14-AUG-1998  (first entry)
DE   Murine macrophage inflammatory protein (MIP)-1 gamma.
KW   Macrophage inflammatory protein; MIP-1 gamma; murine; pyrexia;
KW   treatment; proliferative cell disorder; T-cell-mediated disease;
KW   prophylaxis; bacterial sepsis.
OS   Mus sp.
PN   US5770402-A.
PD   23-JUN-1998.
PF   05-APR-1995; 418032.
PR   05-APR-1995; US-418032.
PA   (TEXA ) UNIV TEXAS SYSTEM.
PI   Beutler BA, Poltorak AN;
DR   WPI; 98-376800/32.
DR   N-PSDB; V34403.
PT   DNA encoding murine macrophage inflammatory protein - useful for
PT   producing recombinant proteins, which can be used to induce pyrexia
PT   in vivo and to treat proliferative cell disorders
PS   Claim 1; Fig 1; 40pp; English.
CC   This represents a murine macrophage inflammatory protein (MIP)-1 gamma
CC   sequence. A recombinant vector containing a MIP-1 gamma encoding DNA
CC   segment, positioned under the control of a promoter can be introduced
CC   into a recombinant host cell. The host cell can be cultured under
CC   conditions effective to allow expression of the encoded MIP-1 gamma
CC   protein. The cells can be used to produce recombinant MIP-1 gamma.
CC   Compositions containing MIP-1 gamma protein or DNA can be used to induce
CC   pyrexia in vivo, to treat proliferative cell disorders and T-cell-
CC   mediated diseases, and for prophylaxis of bacterial sepsis.
SQ   Sequence   122 AA;
```

```
  Query Match             13.8%;  Score 98;  DB 1;  Length 122;
  Best Local Similarity   30.9%;  Pred. No. 6.6e-05;
```

```
  Matches   34;  Conservative   11;  Mismatches   19;  Indels   46;  Gaps    6;

Qy      14 LGAWAPAVHA------------QGV--------FE---DCCLAYHYPIGWAVLRRAWTY 49
           || ||   ||             ||:         |:   ||||:|:
Db      17 LGIWAQITHATETKEVQSSLKAQQGLEIEMFHMGFQDSSDCCLSYN------------S 63

Qy      50 RIQ--------EVSGSCNLPAAIFYLPKRHRKVCGNPKSREVQRAMKLLD 91
           |||          ||:|  |: || ::||   ||:||  | ||| :  |:
Db      64 RIQCSRFIGYFPTSGGCTRPGIIF-ISKRGFQVCANPSDRRVQRCIERLE 112


Search completed: July 24, 1999, 12:41:42
Job time: 4403 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: July 24, 1999, 10:27:23 ; Search time 16.54 Seconds
(without alignments)
363.351 Million cell updates/sec

Title: US-09-002-485-25
Perfect score: 710
Sequence: 1 MNLWLLACLVAGFLGAWAPA.....NPISSSKRNVSLLISANSGL 150

Scoring table: PAM150

Searched: 122810 seqs, 40065486 residues

Database : PIR_60:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	81	11.4	116	2	I49555	gene C10 protein -
2	80	11.3	92	2	I52322	macrophage inflamm
3	79.5	11.2	97	2	JC4912	eotaxin precursor
4	79.5	11.2	120	2	JE0177	lymphocyte and mon
5	76.5	10.8	91	1	A28815	monocyte chemoattr
6	73	10.3	92	2	A32393	macrophage inflamm
7	72.5	10.2	114	1	ETMSL	lymphotactin precu
8	71	10.0	92	2	C30552	macrophage inflamm
9	70	9.9	96	2	A37236	I-309 protein prec
10	69.5	9.8	120	2	I48147	monocyte chemoattr
11	66.5	9.4	91	1	A46539	monocyte chemoattr
12	66	9.3	99	2	JC2417	monocyte chemoattr
13	65.5	9.2	92	2	I46730	immune activation
14	65.5	9.2	99	2	JC5295	monocyte chemotact
15	65	9.2	92	1	A31767	macrophage inflamm
16	65	9.2	96	2	I48099	eotaxin precursor
17	64.5	9.1	92	2	A30574	macrophage inflamm
18	64.5	9.1	608	2	S09790	hypothetical prote
19	63.5	8.9	114	1	ETHUL	lymphotactin precu
20	63	8.9	148	2	A30209	PDGF-inducible JE
21	62.5	8.8	246	1	R5DO7	ribosomal protein
22	62.5	8.8	93	2	B35673	LD78-beta protein
23	62.5	8.8	92	2	S24236	TCA3 protein - mou
24	62.5	8.8	473	2	JC5835	anaphylatoxin C3a
25	62.5	8.8	255	2	A71351	probable cobalt AB
26	62	8.7	622	2	A26981	regulatory protein
27	61.5	8.7	386	2	S34043	oxytocin receptor
28	61.5	8.7	389	2	A55493	oxytocin receptor
29	61.5	8.7	388	2	A55597	oxytocin receptor
30	61.5	8.7	370	2	I48231	serotonin receptor
31	61.5	8.7	443	2	JC4088	chorionic gonadotr
32	61	8.6	302	2	A69431	pyruvate formate-l
33	61	8.6	96	2	JC2478	eotaxin precursor
34	61	8.6	457	2	S08162	nicotinic acetylch
35	60.5	8.5	164	2	S48187	grain softness pro
36	60.5	8.5	567	2	A44225	transforming growt
37	60.5	8.5	534	2	S55635	DNA helicase-prima
38	60	8.5	320	2	S69547	transcription init
39	60	8.5	125	2	I46857	monocyte chemoattr
40	60	8.5	151	2	E71547	hypothetical prote
41	60	8.5	1389	2	T03273	embryogenesis tran
42	60	8.5	1811	3	T00035	nonstructural poly
43	59.5	8.4	429	1	VHVUSR	nucleocapsid prote
44	59.5	8.4	146	2	S67981	hemoglobin beta-I
45	59.5	8.4	916	2	JT0396	reverse transcript

ALIGNMENTS

RESULT 1
I49555
gene C10 protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Aug-1998
C;Accession: I49555
R;Orlofsky, A.; Berger, M.S.; Prystowsky, M.B.
Cell Regul. 2, 403-412, 1991
A;Title: Novel expression pattern of a new member of the MIP-1 family of cytokine-lik
A;Reference number: I49555; MUID:91370083
A;Accession: I49555
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-116 <RES>
A;Cross-references: GB:M58004; NID:g192243; PID:g192244
C;Genetics:
A;Gene: C10
C;Superfamily: macrophage inflammatory protein

Query Match 11.4%; Score 81; DB 2; Length 116;
Best Local Similarity 26.7%; Pred. No. 0.022;
Matches 23; Conservative 13; Mismatches 20; Indels 30; Gaps 4;

```
Qy      18 APAVHA--QGVFEDCCLA-----------YHYPIGWAVLRRAWTYRIQEVSGSCNLPAAI 64
           :| :|   |    |||::           | :|                ||:|  |: |
Db      36 PPIIHQGFQDTSSDCCFSYATQIPCKRFIYYFP---------------TSGGCIKPGII 79

Qy      65 FYLPKRHRKVCGNPKSREVQRAMKLL 90
           | :::|   ||::|  | ||| :  |
Db      80 F-ISRRGTQVCADPSDRRVQRCLSTL 104
```

RESULT 2
I52322
macrophage inflammatory protein-1alpha - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 02-Jul-1998
C;Accession: I52322
R;Shi, M.M.; Godleski, J.J.; Paulauskis, J.D.
Biochem. Biophys. Res. Commun. 211, 289-295, 1995
A;Title: Molecular cloning and posttranscriptional regulation of macrophage inflammat
A;Reference number: I52322; MUID:95298037
A;Accession: I52322
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-92 <RES>
A;Cross-references: EMBL:U22414; NID:g790632; PID:g790633
C;Superfamily: macrophage inflammatory protein

Query Match 11.3%; Score 80; DB 2; Length 92;
Best Local Similarity 27.2%; Pred. No. 0.023;
Matches 22; Conservative 19; Mismatches 20; Indels 20; Gaps 6;

```
Qy       1 MNLWLLACLVAGFLGAWAPAVHAQGVFEDCCLAYHYPIGWAVLRRAWTYRIQ--EVSGSC 58
           | ||      | :: ::   :      ||::|    |: : |:   :  :  | |: |
Db      15 MALW-----NEVFSAPYGADTPT-----ACCFSY----GRQIPRK---FIADYFETSSLC 57

Qy      59 NLPAAIFYLPKRHRKVCGNPK 79
           : |: || | ||:| :|::||
Db      58 SQPGVIF-LTKRNRQICADPK 77
```

```
RESULT   3
JC4912
eotaxin precursor - human
C;Species: Homo sapiens (man)
C;Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 13-Nov-1998
C;Accession: JC4912
R;Bartels, J.; Schlueter, C.; Richter, E.; Noso, N.; Kulke, R.; Christophers, E.; Schroe
Biochem. Biophys. Res. Commun. 225, 1045-1051, 1996
A;Title: Human dermal fibroblasts express eotaxin: Molecular cloning, mRNA expression, a
A;Reference number: JC4912
A;Accession: JC4912
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-97 <BAR>
A;Cross-references: EMBL:Z75668; NID:g1531982; PID:e251275; PID:g1531983
A;Experimental source: dermal fibroblast
C;Comment: This protein has eosinophil specific chemotatic activity.
C;Superfamily: macrophage inflammatory protein
C;Keywords: fibroblast
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-97/Product: eotaxin #status predicted <MAT>


  Query Match             11.2%;  Score 79.5;  DB 2;  Length 97;
  Best Local Similarity   34.0%;  Pred. No. 0.027;
  Matches   32;  Conservative   18;  Mismatches   33;  Indels   11;  Gaps    6;

Qy       3 LWLLACLVAGFLGAWAPA--VHAQGVFEDCCLAYHYPIGWAVLRRAWTYRIQEVSGSCNL 60
           ||||  |:|   :| :|     : :|   ||  :: :     |:|  :|| :  || |
Db       8 LWLL--LIA---AAFSPQGLTGPASVPTTCC--FNLANRKIPLQRLESYR-RITSGKCPQ 59

Qy      61 PAAIFYLPKRHRKVCGNPKSREVQRAMKLLDARN 94
           | ||   |  : :|::|| | || :|| || ::
Db      60 KAVIF-KTKLAKDICADPKKRWVQDSMKYLDQKS 92


RESULT   4
JE0177
lymphocyte and monocyte chemoattractant CC chemokine - human
C;Species: Homo sapiens (man)
C;Date: 10-Jul-1998 #sequence_revision 10-Jul-1998 #text_change 17-Mar-1999
C;Accession: JE0177
R;Youn, B.S.; Zhang, S.; Broxmeryer, H.E.; Antol, K.; Fraser Jr., M.J.; Hangoc, G.; Kwon
Biochem. Biophys. Res. Commun. 247, 217-222, 1998
A;Title: Isolation and characterization of LMC, a novel lymphocyte and monocyte chemoatt
A;Reference number: JE0177; MUID:98308096
A;Accession: JE0177
A;Molecule type: mRNA
A;Residues: 1-120 <YOU>


  Query Match             11.2%;  Score 79.5;  DB 2;  Length 120;
  Best Local Similarity   39.4%;  Pred. No. 0.034;
  Matches   28;  Conservative   10;  Mismatches   20;  Indels   13;  Gaps    5;

Qy      30 CCLAYHYPIGWAVLRR-AWTYRIQEVSGSCNLPAAIFYLPKRHRKVCGNPKSREVQRAMK 88
           ||| |   :    |||    ||    : :|:||| || : ||:| || ||:   ||  :|
Db      37 CCLKYYEKVP---LRRLVVGYRK---ALNCHLPAIIF-VTKRNREVCTNPNDDWVQEYIK 89

Qy      89 -----LLDARN 94
                |: :||
Db      90 DPNLPLIPTRN 100


RESULT   5
A28815
monocyte chemoattractant cytokine RANTES precursor - human
N;Alternate names: small inducible cytokine A5; T-cell specific cytokine RANTES
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1989 #sequence_revision 16-Aug-1996 #text_change 29-May-1998
C;Accession: A28815
R;Schall, T.J.; Jongstra, J.; Dyer, B.J.; Jorgensen, J.; Clayberger, C.; Davis, M.M.;
J. Immunol. 141, 1018-1025, 1988
A;Title: A human T cell-specific molecule is a member of a new gene family.
A;Reference number: A28815; MUID:88285659
A;Accession: A28815
A;Molecule type: mRNA
A;Residues: 1-91 <SCH>
A;Cross-references: GB:M21121
C;Comment: The acronym RANTES reflects the description "Regulated upon Activation, No
C;Genetics:
A;Gene: GDB:SCYA5; D17S136E
A;Cross-references: GDB:120749; OMIM:187011
A;Map position: 17q11.2-17q12
C;Superfamily: macrophage inflammatory protein
C;Keywords: chemotaxis; cytokine; immediate-early protein; inflammation; T-cell
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-91/Product: T-cell protein RANTES #status predicted <MAT>


  Query Match             10.8%;  Score 76.5;  DB 1;  Length 91;
  Best Local Similarity   35.4%;  Pred. No. 0.057;
  Matches   23;  Conservative   17;  Mismatches   14;  Indels   11;  Gaps    5;

Qy      30 CCLAYHYPIGWAVLRRAWTYRIQE---VSGSCNLPAAIFYLPKRHRKVCGNPKSREVQRA 86
           ||:||   |::: | ||   :| |    || |: || :| : :::| ||:||  : |:
Db      33 CCFAY---IARP-LPRA---HIKEYFYTSGKCSNPAVVF-VTRKNRQVCANPEKKWVREY 84

Qy      87 MKLLD 91
           :: |:
Db      85 INSLE 89


RESULT   6
A32393
macrophage inflammatory protein-1-alpha precursor - mouse
N;Alternate names: heparin-binding chemotaxis protein; L2G25B protein; SCI/MIP-1a; SI
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 08-Sep-1997
C;Accession: S11685; A32393; S04533; A53885; A30552; PS0303; A27596; I56104
R;Grove, M.; Lowe, S.; Graham, G.; Pragnell, I.; Plumb, M.
Nucleic Acids Res. 18, 5561, 1990
A;Title: Sequence of the murine haemopoietic stem cell inhibitor/macrophage inflammat
A;Reference number: S11685; MUID:91016858
A;Accession: S11685
A;Molecule type: DNA
A;Residues: 1-92 <GRO>
A;Cross-references: EMBL:X53372; NID:g54062; PID:g297531
A;Note: the authors' translation of the nucleotide sequence differs at several positi
R;Kwon, B.S.; Weissman, S.M.
Proc. Natl. Acad. Sci. U.S.A. 86, 1963-1967, 1989
A;Title: cDNA sequence of two inducible T-cell genes.
A;Reference number: A32393; MUID:89184547
A;Accession: A32393
A;Molecule type: mRNA
A;Residues: 1-92 <KWO>
A;Cross-references: GB:J04491; NID:g201524; PID:g201525
R;Davatelis, G.; Tekamp-Olson, P.; Wolpe, S.D.; Hermsen, K.; Luedke, C.; Gallegos, C.
J. Exp. Med. 167, 1939-1944, 1988
A;Title: Cloning and characterization of a cDNA for murine macrophage inflammatory pr
A;Reference number: S04533; MUID:88258380
A;Accession: S04533
A;Molecule type: mRNA
A;Residues: 1-48,'E',50-90,'I',92 <DA2>
A;Cross-references: EMBL:X12531
A;Note: the authors translated the codon GAG for residue 49 as Asp and ATT for residu
A;Note: the sequence has been corrected in reference A53885
R;Davatelis, G.; Tekamp-Olson, P.; Wolpe, S.D.; Hermsen, K.; Luedke, C.; Gallegos, C.
J. Exp. Med. 170, 2189, 1989
A;Reference number: A53885
A;Contents: erratum
A;Accession: A53885
A;Molecule type: mRNA
A;Residues: 1-92 <DAV>
```

```
A;Cross-references: EMBL:X12531; NID:g53122; PID:g53123
R;Brown, K.D.; Zurawski, S.M.; Mosmann, T.R.; Zurawski, G.
J. Immunol. 142, 679-687, 1989
A;Title: A family of small inducible proteins secreted by leukocytes are members of a ne
s of various activation processes.
A;Reference number: A30552; MUID:89093958
A;Accession: A30552
A;Molecule type: mRNA
A;Residues: 1-21,'L',23-61,'A',63-92 <BRO>
A;Cross-references: GB:M23447; NID:g533240; PID:g533241
R;Sherry, B.; Tekamp-Olson, P.; Gallegos, C.; Bauer, D.; Davatelis, G.; Wolpe, S.D.; Mas
J. Exp. Med. 168, 2251-2259, 1988
A;Title: Resolution of the two components of macrophage inflammatory protein 1, and clon
A;Reference number: JL0088; MUID:89067830
A;Accession: PS0303
A;Molecule type: mRNA
A;Residues: 24-33,'XX',36-54 <SHE>
R;Wolpe, S.D.; Davatelis, G.; Sherry, B.; Beutler, B.; Hesse, D.G.; Nguyen, H.T.; Moldaw
J. Exp. Med. 167, 570-581, 1988
A;Title: Macrophages secrete a novel heparin-binding protein with inflammatory and neutr
A;Reference number: A27596; MUID:88154745
A;Accession: A27596
A;Molecule type: protein
A;Residues: 24-33,'XX',36-42 <WOL>
A;Note: 26-Met, 30-Pro, and 39-Thr were also found
R;Widmer, U.; Yang, Z.; van Deventer, S.; Manogue, K.R.; Sherry, B.; Cerami, A.
J. Immunol. 146, 4031-4040, 1991
A;Title: Genomic structure of murine macrophage inflammatory protein-1-alpha and conserv
A;Reference number: I56104; MUID:91237116
A;Accession: I56104
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-92 <RES>
A;Cross-references: GB:M73061; NID:g199694; PID:g199695
C;Comment: This protein is a monokine.
C;Genetics:
A;Introns: 23/3; 26/1; 63/2
C;Superfamily: macrophage inflammatory protein
C;Keywords: heparin binding
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-92/Product: macrophage inflammatory protein #status experimental <MAT>


  Query Match             10.3%;  Score 73;  DB 2;  Length 92;
  Best Local Similarity   30.8%;  Pred. No. 0.15;
  Matches   16;  Conservative   15;  Mismatches   11;  Indels   10;  Gaps    3;

Qy      30 CCLAYHYPIGWAVLRRAWTYRIQ--EVSGSCNLPAAIFYLPKRHRKVCGNPK 79
           ||::|    |        :: : ::  | |: |: |: || | ||:| :|:::|
Db      34 CCFSYSRKI-------PRQFIVDYFETSSLCSQPGVIF-LTKRNRQICADSK 77


RESULT   7
ETMSL
lymphotactin precursor - mouse
N;Alternate names: single cysteine motif-1, SCM-1
C;Species: Mus musculus (house mouse)
C;Date: 23-Mar-1995 #sequence_revision 07-Jun-1996 #text_change 26-Feb-1999
C;Accession: A55248; I67635
R;Kelner, G.S.; Kennedy, J.; Bacon, K.B.; Kleyensteuber, S.; Largaespada, D.A.; Jenkins,
Science 266, 1395-1399, 1994
A;Title: Lymphotactin: a cytokine that represents a new class of chemokine.
A;Reference number: A55248; MUID:95064019
A;Accession: A55248
A;Molecule type: mRNA
A;Residues: 1-114 <KEL>
A;Cross-references: GB:U15607; NID:g595908; PID:g595909
A;Note: authors translated the codon AAA for residue 98 as Leu, GTA for residue 110 as I
R;Yoshida, T.; Imai, T.; Kakizaki, M.; Nishimura, M.; Yoshie, O.
FEBS Lett. 360, 155-159, 1995
A;Title: Molecular cloning of a novel C or gamma type chemokine, SCM-1.
A;Reference number: I53506; MUID:95180438
A;Accession: I67635
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-109,'I',111-114 <RES>
A;Cross-references: GB:D43769; NID:g927656; PID:g927657
C;Comment: Lymphotactin is produced by activated T-cells and is a chemotactin for mon
C;Superfamily: lymphotactin
C;Keywords: chemotaxis; cytokine; lymphokine; mast cell; T-cell
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-21/Domain: propeptide #status predicted <PRO>
F;22-114/Product: lymphotactin #status predicted <MAT>


  Query Match             10.2%;  Score 72.5;  DB 1;  Length 114;
  Best Local Similarity   23.9%;  Pred. No. 0.21;
  Matches   27;  Conservative   31;  Mismatches   44;  Indels   11;  Gaps    4;

Qy       1 MNLWLLACLVAGFLGAWAPAVHAQGVFED-CCLAYHYPIGWAVLRRAWTYRIQEVSGSCN 59
           | | ||: |    | :|      :  |:|:  |:  :    :  :::  || | | :
Db       1 MRLLLLTFLGVCCLTPWVVEGVGTEVLEESSCVNLQ--TQRLPVQKIKTYIIWEGAMR-- 57

Qy      60 LPAAIFYLPKRHRKVCGNPKSREVQRAMKLLDARNKVFAKLRHNTQTFQAGPH 112
              |: :: ||  |:|::| :: |  |:| :|:|    :  :: ::|  :|::
Db      57 ---AVIFVTKRGLKICADPEAKWVKAAIKTVDGRA---STRKNMAETVPTGAQ 103


RESULT   8
C30552
macrophage inflammatory protein 1-beta precursor - mouse
N;Alternate names: H400; SIS gamma; T-cell activation protein gamma
C;Species: Mus musculus (house mouse)
C;Date: 28-Aug-1989 #sequence_revision 28-Aug-1989 #text_change 13-Nov-1998
C;Accession: C30552; JL0088; PS0304; S22042
R;Brown, K.D.; Zurawski, S.M.; Mosmann, T.R.; Zurawski, G.
J. Immunol. 142, 679-687, 1989
A;Title: A family of small inducible proteins secreted by leukocytes are members of a
s of various activation processes.
A;Reference number: A30552; MUID:89093958
A;Accession: C30552
A;Molecule type: mRNA
A;Residues: 1-92 <BRO>
A;Cross-references: GB:M23503; NID:g533244; PID:g533245
R;Sherry, B.; Tekamp-Olson, P.; Gallegos, C.; Bauer, D.; Davatelis, G.; Wolpe, S.D.;
J. Exp. Med. 168, 2251-2259, 1988
A;Title: Resolution of the two components of macrophage inflammatory protein 1, and c
A;Reference number: JL0088; MUID:89067830
A;Accession: JL0088
A;Molecule type: mRNA
A;Residues: 1-92 <SHE>
A;Cross-references: GB:M35590; NID:g199696; PID:g199697
A;Accession: PS0304
A;Molecule type: protein
A;Residues: 24-33,'XX',36,'X',38 <SH2>
R;Daubersies, P.; Lepretre, F.; Bailleul, B.; Grove, M.; Pragnell, I.; Plumb, M.
submitted to the EMBL Data Library, October 1991
A;Description: Sequence of the murine macrophage inflammatory protein 1b gene.
A;Reference number: S22042
A;Accession: S22042
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-92 <DAU>
A;Cross-references: EMBL:X62502; NID:g53126; PID:g53127
C;Comment: This protein is a monokine.
C;Genetics:
A;Introns: 26/1; 64/2
C;Superfamily: macrophage inflammatory protein
C;Keywords: glycoprotein
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-92/Product: macrophage inflammatory protein 1-beta #status experimental <MAT>
F;76/Binding site: carbohydrate (Asn) (covalent) #status predicted


  Query Match             10.0%;  Score 71;  DB 2;  Length 92;
  Best Local Similarity   32.1%;  Pred. No. 0.25;
```

```
  Matches   27;  Conservative   14;  Mismatches   21;  Indels   22;  Gaps    5;

Qy      3 LWLLACLVAGF--------LGAWAPAVHAQGVFEDCCLAYHYPIGWAVLRRAWTYRIQEV 54
          | ||  |||:|        :|: :|:         ||:::|        |:|:      |
Db      8 LSLLL-LVAAFCAPGFSAPMGSDPPT--------SCCFSY----TSRQLHRSFVMDYYET 54

Qy     55 SGSCNLPAAIFYLPKRHRKVCGNP 78
          |: |: || :| | || | :|:||
Db     55 SSLCSKPAVVF-LTKRGRQICANP 77


RESULT   9
A37236
I-309 protein precursor - human
C;Species: Homo sapiens (man)
C;Date: 30-Dec-1991 #sequence_revision 30-Dec-1991 #text_change 24-Sep-1998
C;Accession: A37236; A45817
R;Miller, M.D.; Wilson, S.D.; Dorf, M.E.; Seuanez, H.N.; O'Brien, S.J.; Krangel, M.S.
J. Immunol. 145, 2737-2744, 1990
A;Title: Sequence and chromosomal location of the I-309 gene. Relationship to genes enco
A;Reference number: A37236; MUID:91010756
A;Accession: A37236
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-96 <MIL>
A;Cross-references: GB:M57506; NID:g184505; PID:g184506
R;Miller, M.D.; Hata, S.; Malefyt, R.D.W.; Krangel, M.S.
J. Immunol. 143, 2907-2916, 1989
A;Title: A novel polypeptide secreted by activated human T lymphocytes.
A;Reference number: A45817; MUID:90038522
A;Accession: A45817
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-96 <MI2>
A;Cross-references: GB:M57502; NID:g339728; PID:g339729
C;Genetics:
A;Gene: GDB:SCYA1; I-309
A;Cross-references: GDB:118872; OMIM:182281
A;Map position: 17q12-17q12
C;Superfamily: macrophage inflammatory protein


  Query Match             9.9%;  Score 70;  DB 2;  Length 96;
  Best Local Similarity   29.7%;  Pred. No. 0.35;
  Matches   27;  Conservative   16;  Mismatches   32;  Indels   16;  Gaps    6;

Qy      6 LACLVAGFLGAWAPAVHAQGV---FEDCCLAY---HYPIGWAVLRRAWTYRIQEVSGSCN 59
          | ||:  : | |:  | : ::   |  ||:::     |     ||    ||  : |: |:
Db      8 LVCLL--LAGMWPEDVDSKSMQVPFSRCCFSFAEQEIP-----LRAILCYR--NTSSICS 58

Qy     60 LPAAIFYLPKRHRKVCGNPKSREVQRAMKLL 90
              : || | || :  |:      |||  |:|
Db     59 NEGLIFKL-KRGKEACALDTVGWVQRHRKML 88


RESULT  10
I48147
monocyte chemoattractant protein-1 - guinea pig
C;Species: Cavia porcellus (guinea pig)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-May-1997
C;Accession: I48147
R;Yoshimura, T.
J. Immunol. 150, 5025-5032, 1993
A;Title: cDNA cloning of guinea pig monocyte chemoattractant protein-1 and expression of
A;Reference number: I48147; MUID:93267104
A;Accession: I48147
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-120 <RES>
A;Cross-references: GB:L04985; NID:g349820; PID:g349821
C;Genetics:
A;Gene: MCP-1
C;Superfamily: macrophage inflammatory protein


  Query Match             9.8%;  Score 69.5;  DB 2;  Length 120;
  Best Local Similarity   25.6%;  Pred. No. 0.5;
  Matches   31;  Conservative   29;  Mismatches   36;  Indels   25;  Gaps    8;

Qy      5 LLACLVA------GFLGAWAPAVHAQGVFEDCCLAYHYPIGWAVLRRAWTY-RIQEVSGS 57
          :| ||:       ::| | : :|::      || :::  |    |:|   | ||   |:
Db      6 VLLCLLVIEATFCSLLMAQPDGVNTP----TCCYTFNKQIP---LKRVKGYERI--TSSR 56

Qy     58 CNLPAAIFYLPKRHRKVCGNPKSREVQRAMKLLDARNKVFAKLRHNTQTFQAGPHAVKKL 117
          |   | ||    ::: ||::|  : ||  :  || |    :  ::|:    ::|:: | |
Db     57 CPQEAVIF-RTLKNKEVCADPTQKWVQDYIAKLDQR----TQQKQNS----TAPQTSKPL 107

Qy    118 S 118
          :
Db    108 N 108


RESULT  11
A46539
monocyte chemoattractant cytokine RANTES precursor - mouse
N;Alternate names: MuRantes
C;Species: Mus musculus (house mouse)
C;Date: 18-Jun-1993 #sequence_revision 16-Aug-1996 #text_change 22-Jan-1999
C;Accession: I48875; A46539; I48654; I56970
R;Danoff, T.M.; Lalley, P.A.; Chang, Y.S.; Heeger, P.S.; Neilson, E.G.
J. Immunol. 152, 1182-1189, 1994
A;Title: Cloning, genomic organization, and chromosomal localization of the Scya5 gen
A;Reference number: I48875; MUID:94132613
A;Accession: I48875
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-91 <DAN>
A;Cross-references: EMBL:U02298; NID:g460090; PID:g460091
R;Schall, T.J.; Simpson, N.J.; Mak, J.Y.
Eur. J. Immunol. 22, 1477-1481, 1992
A;Title: Molecular cloning and expression of the murine RANTES cytokine: structural a
A;Reference number: A46539; MUID:92289805
A;Accession: A46539
A;Molecule type: mRNA
A;Residues: 1-18,'A',20-91 <SCH>
A;Cross-references: GB:S37648; NID:g250207; PID:g250208
A;Experimental source: macrophage cell line PU5-1.8
A;Note: sequence extracted from NCBI backbone (NCBIN:106768, NCBIP:106770)
R;Shin, H.S.; Drysdale, B.E.; Shin, M.L.; Noble, P.W.; Fisher, S.N.; Paznekas, W.A.
Mol. Cell. Biol. 14, 2914-2925, 1994
A;Title: Definition of a lipopolysaccharide-responsive element in the 5'-flanking reg
A;Reference number: I48654; MUID:94217689
A;Accession: I48654
A;Status: translation not shown; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-91 <SHI>
A;Cross-references: EMBL:X70675; NID:g475205; PID:g475206
R;Neilson, E.G.; Krensky, A.
Kidney Int. 41, 220-225, 1992
A;Title: Isolation and characterization of cDNA from renal tubular epithelium encodin
A;Reference number: I56970; MUID:92277990
A;Accession: I56970
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-40,'E',42-91 <NEI>
A;Cross-references: GB:M77747; NID:g200649; PID:g200650
C;Comment: This chemoattractant for monocytes but not neutrophils is an immediate-ear
C;Genetics:
A;Introns: 26/1; 63/2
C;Superfamily: macrophage inflammatory protein
C;Keywords: chemotaxis; cytokine; immediate-early protein; inflammation
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-91/Product: monocyte chemoattractant cytokine RANTES #status predicted <MAT>


  Query Match             9.4%;  Score 66.5;  DB 1;  Length 91;
```

```
  Best Local Similarity   32.3%;  Pred. No. 0.84;
  Matches   21;  Conservative   14;  Mismatches   19;  Indels   11;  Gaps    4;

Qy      30 CCLAYHYPIGWAVLRRAWTYRIQE---VSGSCNLPAAIFYLPKRHRKVCGNPKSREVQRA 86
           ||:||       | ||   :: |    |: |:  | :| : :|:| ||:||  : ||
Db      33 CCFAYL----SLALPRA---HVKEYFYTSSKCSNLAVVF-VTRRNRQVCANPEKKWVQEY 84

Qy      87 MKLLD 91
           :: |:
Db      85 INYLE 89


RESULT  12
JC2417
monocyte chemoattractant protein-2 precursor - pig
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 24-Feb-1995 #sequence_revision 24-Feb-1995 #text_change 17-Mar-1999
C;Accession: JC2417
R;Hosang, K.; Knoke, I.; Klaudiny, J.; Wempe, F.; Wuttke, W.; Scheit, K.H.
Biochem. Biophys. Res. Commun. 205, 148-153, 1994
A;Title: Porcine luteal cells express monocyte chemoattractant protein-2 (MCP-2): Analys
A;Reference number: JC2417; MUID:95091716
A;Accession: JC2417
A;Molecule type: mRNA
A;Residues: 1-99 <HOS>
A;Experimental source: corpus luteum
C;Superfamily: macrophage inflammatory protein
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-99/Product: monocyte chemoattractant protein-2 #status predicted <MAT>


  Query Match             9.3%;  Score 66;  DB 2;  Length 99;
  Best Local Similarity   37.5%;  Pred. No. 1.1;
  Matches   12;  Conservative    9;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      63 AIFYLPKRHRKVCGNPKSREVQRAMKLLDARN 94
           |: : |  : ||::|  : || :|||| ::
Db      63 AVIFKTKADKEVCADPQQKWVQNSMKLLDQKS 94


RESULT  13
I46730
immune activation gene 2 - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 14-Feb-1997 #sequence_revision 14-Feb-1997 #text_change 09-May-1997
C;Accession: I46730
R;Mori, S.; Goto, K.; Goto, F.; Mutakami, K.; Ohkawara, S.; Yoshinaga, M.
Int. Immunol. 6, 149-156, 1994
A;Title: Dynamic changes in mRNA expression of neutrophils during the course of acute in
A;Reference number: I46730; MUID:94198229
A;Accession: I46730
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-92 <MOR>
A;Cross-references: GB:D17402; NID:g599577; PID:g599578
C;Superfamily: macrophage inflammatory protein


  Query Match             9.2%;  Score 65.5;  DB 2;  Length 92;
  Best Local Similarity   32.7%;  Pred. No. 1.1;
  Matches   17;  Conservative   13;  Mismatches   11;  Indels   11;  Gaps    4;

Qy      30 CCLAYHYPIGWAVLRR-AWTYRIQ--EVSGSCNLPAAIFYLPKRHRKVCGNP 78
           ||::|       ||: :: : |:  | :: |: || :|   |: | ||:||
Db      34 CCFSY-------TLRKLPRHFVIDYFETTSLCSQPAVVF-QTKKGRQVCANP 77


RESULT  14
JC5295
monocyte chemotactic protein-2 precursor - human
C;Species: Homo sapiens (man)
C;Date: 02-May-1997 #sequence_revision 18-Jul-1997 #text_change 17-Mar-1999
C;Accession: JC5295
R;Van Coillie, E.; Froyen, G.; Nomiyama, H.; Miura, R.; Fiten, P.; Van Aelst, I.; Van
Biochem. Biophys. Res. Commun. 231, 726-730, 1997
A;Title: Human monocyte chemotactic protein-2: cDNA cloning and regulated expression
A;Reference number: JC5295; MUID:97224420
A;Accession: JC5295
A;Molecule type: mRNA
A;Residues: 1-99 <VAN>
A;Cross-references: GB:Y10802; NID:g1924937; PID:e294088; PID:g1924938
A;Experimental source: bone marrow
C;Comment: This protein belongs to the beta-chemokine family which is one of the majo
tis and in tumor biology, and contribute to the trafficking and recuitment of the res
C;Genetics:
A;Gene: mcp-2
C;Superfamily: macrophage inflammatory protein
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-99/Product: monocyte chemotactic protein-2 #status predicted <MAT>


  Query Match             9.2%;  Score 65.5;  DB 2;  Length 99;
  Best Local Similarity   35.9%;  Pred. No. 1.2;
  Matches   14;  Conservative   10;  Mismatches   12;  Indels    3;  Gaps    1;

Qy      63 AIFYLPKRHRKVCGNPKSREVQRAMKLLDARNKVFAKLR 101
           |: :   | : ||::|| | |: :|| ||    :| :|:
Db      63 AVIFKTQRGKEVCADPKERWVRDSMKHLD---QIFQNLK 98


RESULT  15
A31767
macrophage inflammatory protein 1-beta precursor - human
N;Alternate names: cytokine HC21; G-26 protein; H400 homolog; lymphocyte activation g
protein 2 (Act-2); T-cell activation protein gamma
C;Species: Homo sapiens (man)
C;Date: 07-Jun-1990 #sequence_revision 29-May-1998 #text_change 29-May-1998
C;Accession: JH0319; A40978; A31767; A37411; B30574; B45817; D30552
R;Baixeras, E.; Roman-Roman, S.; Jitsukawa, S.; Genevee, C.; Mechiche, S.; Viegas-Peq
Mol. Immunol. 27, 1091-1102, 1990
A;Title: Cloning and expression of a lymphocyte activation gene (LAG-1).
A;Reference number: JH0319; MUID:91061800
A;Accession: JH0319
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-92 <BAI>
A;Cross-references: GB:X53682; NID:g34217; PID:g34218
A;Experimental source: natural killer cell, strain CD3-CD2+, F5, 5IIIE5
R;Napolitano, M.; Modi, W.S.; Cevario, S.J.; Gnarra, J.R.; Seuanez, H.N.; Leonard, W.
J. Biol. Chem. 266, 17531-17536, 1991
A;Title: The gene encoding the Act-2 cytokine. Genomic structure, HTLV-I/tax responsi
A;Reference number: A40978; MUID:91373378
A;Accession: A40978
A;Molecule type: DNA
A;Residues: 1-14,'S',16-69,'G',71-92 <NAP>
A;Cross-references: GB:M69201; NID:g178021
A;Note: 15-Ala was also found
R;Lipes, M.A.; Napolitano, M.; Jeang, K.T.; Chang, N.T.; Leonard, W.J.
Proc. Natl. Acad. Sci. U.S.A. 85, 9704-9708, 1988
A;Title: Identification, cloning, and characterization of an immune activation gene.
A;Reference number: A31767; MUID:89071764
A;Accession: A31767
A;Molecule type: mRNA
A;Residues: 1-92 <LIP>
A;Cross-references: GB:J04130; NID:g178017; PID:g178018
R;Chang, H.C.; Reinherz, E.L.
Eur. J. Immunol. 19, 1045-1051, 1989
A;Title: Isolation and characterization of a cDNA encoding a putative cytokine which
A;Reference number: A37411; MUID:89325421
A;Accession: A37411
A;Molecule type: mRNA
A;Residues: 1-92 <CHA>
A;Cross-references: GB:X16166; NID:g32035; PID:g32036
R;Zipfel, P.F.; Balke, J.; Irving, S.G.; Kelly, K.; Siebenlist, U.
J. Immunol. 142, 1582-1590, 1989
A;Title: Mitogenic activation of human T cells induces two closely related genes whic
```

```
A;Reference number: A30574; MUID:89140347
A;Accession: B30574
A;Molecule type: mRNA
A;Residues: 1-19,'L',21-92 <ZIP>
A;Cross-references: GB:M25316; NID:g602454; PID:g602455
R;Miller, M.D.; Hata, S.; Malefyt, R.D.W.; Krangel, M.S.
J. Immunol. 143, 2907-2916, 1989
A:Title: A novel polypeptide secreted by activated human T lymphocytes.
A;Reference number: A45817; MUID:90038522
A;Accession: B45817
A;Molecule type: mRNA
A;Residues: 7-55,'I',57-79,'T',81-92 <MIL>
A;Cross-references: GB:M57503; NID:g339726; PID:g339727
R;Brown, K.D.; Zurawski, S.M.; Mosmann, T.R.; Zurawski, G.
J. Immunol. 142, 679-687, 1989
A;Title: A family of small inducible proteins secreted by leukocytes are members of a ne
s of various activation processes.
A;Reference number: A30552; MUID:89093958
A;Accession: D30552
A;Molecule type: mRNA
A;Residues: 1-39,'REASS',46-92 <BRO>
A;Cross-references: GB:M23502; NID:g533212; PID:g533213
R;Clore, G.M.; Lodi, P.J.; Garrett, D.S.; Gronenborn, A.M.
submitted to the Brookhaven Protein Data Bank, January 1994
A;Reference number: A52206; PDB:1HUM
A;Contents: annotation; conformation and disulfide bond assignments by (1)H-NMR, residue
C;Comment: This protein is secreted by activated lymphocytes and monocytes. It is bound
C;Genetics:
A;Gene: GDB:LAG1
A;Cross-references: GDB:127451; OMIM:153335
A;Map position: 17q21-17q21
A;Introns: 26/1; 64/2
C;Superfamily: macrophage inflammatory protein
C;Keywords: chemotaxis; cytokine; inflammation
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-92/Product: macrophage inflammatory protein 1-beta #status experimental <MAT>
F;34-58,35-74/Disulfide bonds: #status experimental


  Query Match             9.2%;  Score 65;  DB 1;  Length 92;
  Best Local Similarity   26.7%;  Pred. No. 1.3;
  Matches   23;  Conservative   20;  Mismatches   17;  Indels   26;  Gaps    6;

Qy       3 LWLLACLVAGF--------LGAWAPAVHAQGVFEDCCLAYHYPIGWAVLRRAWTYRIQ-- 53
           | ||  |||:|        :|: :|:         ||::|      :  : :: : ::
Db       8 LSLLM-LVAAFCSPALSAPMGSDPPT--------ACCFSY------TARKLPRNFVVDYY 52

Qy      53 EVSGSCNLPAAIFYLPKRHRKVCGNP 78
           | |: |: || :|   || : ||::|
Db      53 ETSSLCSQPAVVF-QTKRSKQVCADP 77



Search completed: July 24, 1999, 12:41:13
Job time: 8030 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 24, 1999, 07:05:07 ; Search time 427.47 Seconds
(without alignments)
7380.323 Million cell updates/sec

Title: US-09-002-485-102
Perfect score: 992
Sequence: 1 GACAGCTTGGCCTACAGCCC.....TGATGGAGTCAAAAAAAAAA 992

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_st:*
14: gb_sts:*
15: gb_sy:*
16: gb_un:*
17: gb_vi:*
18: em_fun:*
19: em_htg:*
20: em_hum1:*
21: em_hum2:*
22: em_in:*
23: em_om:*
24: em_or:*
25: em_ov:*
26: em_pat:*
27: em_ph:*
28: em_pl:*
29: em_ro:*
30: em_sts:*
31: em_sy:*
32: em_un:*
33: em_vi:*
34: gb_htg1:*
35: gb_htg2:*
36: gb_in1:*
37: gb_in2:*
38: em_ba1:*
39: em_ba2:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	807.6	81.4	879	11	HSU86358	U86358 Human chemo
2	130.6	13.2	892	12	MMU86357	U86357 Mus musculu
3	51.2	5.2	7218	5	I66494	I66494 Sequence 14
c 4	46.2	4.7	2451	3	AF046861	AF046861 Bos tauru
5	42.8	4.3	40198	10	HSL261H12	Z54072 Human DNA s
c 6	41	4.1	8200	1	STMDNAGYR	L27063 Streptomyce
7	40.2	4.1	345	9	HSAJ3461	AJ223461 Homo sapi
8	39.4	4.0	22550	1	MTCY22G8	Z95585 Mycobacteri
9	38.6	3.9	360	10	HSU85768	U85768 Human myelo
10	37.6	3.8	67200	1	MTV017	AL021897 Mycobacte
11	37.6	3.8	22789	1	SC2A11	AL031184 Streptomy
c 12	37.6	3.8	199248	35	AC007387	AC007387 Homo sapi
13	37.4	3.8	4702	3	CFY15483	Y15483 Canis famil
14	37.4	3.8	14515	3	CFY15484	Y15484 Canis famil
c 15	37.4	3.8	603	8	AF019858	AF019858 Gynerium
16	37.4	3.8	1473	8	AF022740	AF022740 Oryza sat
17	37.4	3.8	1632	17	BHV5GC	Z49224 Bovine herp
c 18	37.2	3.8	36583	1	SC5H1	AL049863 Streptomy
19	37	3.7	23550	1	SCE39	AL049573 Streptomy
c 20	36.8	3.7	1754	7	ZMTNMU17	Y00603 Zea mays tr
21	36.8	3.7	7371	12	AB004329	AB004329 Rattus no
22	36.6	3.7	37730	1	SCE9	AL049841 Streptomy
c 23	36.6	3.7	3905	10	HSHB9HB1	U07663 Human HB9 h
c 24	36.4	3.7	2975	1	MGNMAGA	D32253 Magnetospir
c 25	36.4	3.7	36734	1	SC6G10	AL049497 Streptomy
c 26	36.4	3.7	40523	10	HSN114B2	Z73416 Human DNA s
27	36.4	3.7	969	10	HUMTCREC	L19936 Human T cel
28	36.4	3.7	232650	11	U66061	U66061 Human germl
c 29	36.2	3.6	37200	1	SCE63	AL035640 Streptomy
c 30	36.2	3.6	5130	2	REU45318	U45318 Rhizobium e
31	36	3.6	13285	1	AMLUNA	L36679 Amycolatops
c 32	36	3.6	171188	34	AC005025	AC005025 Homo sapi
c 33	35.8	3.6	109646	11	HSMHCT8S22	AF019413 Homo sapi
c 34	35.8	3.6	129811	34	HSDJ34F7	AL049547 Homo sapi
c 35	35.6	3.6	1218	1	FVBPOAD2	D10686 Flavobacter
36	35.6	3.6	45519	1	FVBPOAD2A	D26094 Flavobacter
c 37	35.6	3.6	2929	2	SCCLAVAM1	AF124928 Streptomy
c 38	35.6	3.6	1068	5	A59400	A59400 Sequence 4
c 39	35.6	3.6	1068	5	A59401	A59401 Sequence 1
c 40	35.6	3.6	159515	12	AC002327	AC002327 Mus muscu
c 41	35.6	3.6	20000	12	AF068865	AF068865 Mus muscu
42	35.6	3.6	1385	12	MMU47281	U47281 Mus musculu
43	35.4	3.6	172048	9	HS179N16	Z95152 Homo sapien
44	35.4	3.6	191356	42	AC005041	AC005041 Homo sapi
45	35.2	3.5	121576	34	AC003117	AC003117 Homo sapi

ALIGNMENTS

RESULT 1
HSU86358
LOCUS HSU86358 879 bp mRNA PRI 11-SEP-1997
DEFINITION Human chemokine (TECK) mRNA, complete cds.
ACCESSION U86358
NID g2388626
VERSION U86358.1 GI:2388626
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 879)
AUTHORS Vicari,A.P., Figueroa,D.J., Hedrick,J.A., Foster,J.S., Singh,K.P., Menon,S., Copeland,N.G., Gilbert,D.J., Jenkins,N.A., Bacon,K.B. and Ziotnik,A.
TITLE TECK: a novel cc chemokine specifically expressed by thymic dendritic cells and potentially involved in T cell development
JOURNAL Immunology 7, 291-301 (1997)
REFERENCE 2 (bases 1 to 879)
AUTHORS Vicari,A.P. and Zlotnik,A.
TITLE Direct Submission
JOURNAL Submitted (21-JAN-1997) Immunology, DNAX Research Institute, 901 California Ave., Palo Alto, CA 94304, USA

```
FEATURES             Location/Qualifiers
     source          1. .879
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /tissue_type="thymus"
     gene            1. .879
                     /gene="TECK"
     CDS             1. .453
                     /gene="TECK"
                     /codon_start=1
                     /product="chemokine"
                     /protein_id="AAB69981.1"
                     /db_xref="PID:g2388627"
                     /db_xref="GI:2388627"
                     /translation="MNLWLLACLVAGFLGAWAPAVHTQGVFEDCCLAYHYPIGWAVLR
                     RAWTYRIQEVSGSCNLPAAIFYLPKRHRKVCGNPKSREVQRAMKLLDARNKVFAKLHH
                     NMQTFQAGPHAVKKLSSGNSKLSSSKFSNPISSSKRNVSLLISANSGL"
BASE COUNT      191 a    264 c    218 g    206 t
ORIGIN
```

Query Match 81.4%; Score 807.6; DB 11; Length 879;
Best Local Similarity 96.9%; Pred. No. 3.9e-193;
Matches 855; Conservative 0; Mismatches 24; Indels 3; Gaps 3;

```
Qy    106 ATGAACCTGTGGCTCCTGGCCTGCCTGGTGGCCGGCTTCCTGGGAGCCTGGGCCCCCGCT 165
Db      1 ATGAACCTGTGGCTCCTGGCCTGCCTGGTGGCCGGCTTCCTGGGAGCCTGGGCCCCCGCT 60

Qy    166 GTCCACGCCCAAGGTGTCTTTGAGGACTGCTGCCTGGCCTACCACTACCCCATTGGGTGG 225
Db     61 GTCCACACCCAAGGTGTCTTTGAGGACTGCTGCCTGGCCTACCACTACCCCATTGGGTGG 120

Qy    226 GCTGTGCTCCGGCGCGCCTGGACTTACCGGATCCAGGAGGTGAGCGGGAGCTGCAATCTG 285
Db    121 GCTGTGCTCCGGCGCGCCTGGACTTACCGGATCCAGGAGGTGAGCGGGAGCTGCAATCTG 180

Qy    286 CCTGCTGCGATATTCTACCTCCCCAAGAGACACAGGAAGGTGTGTGGGAACCCCAAAAGC 345
Db    181 CCTGCTGCGATATTCTACCTCCCCAAGAGACACAGGAAGGTGTGTGGGAACCCCAAAAGC 240

Qy    346 AGGGAGGTGCAGAGAGCCATGAAGCTCCTGGATGCTCGAAATAAGGTTTTTGCAAAGCTC 405
Db    241 AGGGAGGTGCAGAGAGCCATGAAGCTCCTGGATGCTCGAAATAAGGTTTTTGCAAAGCTC 300

Qy    406 CGCCACAACACGCAGACCTTCCAAGCAGGCCCTCATGCTGTAAAGAAGTTGAGTTCTGGA 465
Db    301 CACCACAACATGCAGACCTTCCAAGCAGGCCCTCATGCTGTAAAGAAGTTGAGTTCTGGA 360

Qy    466 AACTCCAAGTTATCATCATCCAAGTTTAGCAATCCCATCAGCAGCAGCAAGAGGAATGTC 525
Db    361 AACTCCAAGTTATCATCATCCAAGTTTAGCAATCCCATCAGCAGCAGCAAGAGGAATGTC 420

Qy    526 TCCCTCCTGATATCAGCTAATTCAGGACTGTGAGCCGGCTCATTTCTGGGCTCCATCGGC 585
Db    421 TCCCTCCTGATATCAGCTAATTCAGGACTGTGAGCCGGCTCATTTCTGGGCTCCATCGGC 480

Qy    586 ACAGGAGGGGCCGGATCTTTCTCCGATAAAACCGTCGCCCTACAGACCCAGCTGTCCCCA 645
Db    481 ACAGGAGGGGCCGGATCTTTCTCCGATAAAACCGTCGCCCTACAGACCCAGCTGTCCCCA 540

Qy    646 CGCCTCTGTCTTTTGGGTCAAGTCTTAATCCCTGCACCTGAGTTGGTCCTCCCTCTGCAC 705
Db    541 CGCCTCTGTCTTTTGGGTCAAGTCTTAATCCCTGCACCTGAGTTGGTCCTCCCTCTGCAC 600

Qy    706 CCCCACCACCTCCTGCCCGTCTGGCAACTGGAAAGAGGGAGTTGGCCTGATTTTAAGCCT 765
Db    601 CCCCACCACCTCCTGCCCGTCTGGCAACTGGAAAGAAGGAGTTGGCCTGATTTTAA-CCT 659

Qy    766 TTTGCCGCTCCGGGGACCAGCAATCCTGGGCAGCCAGTGGCTCTTGTAGAGAAGACT 825
Db    660 TTTGCCGCTCCGGGGAACAGCA-CAATCCTGGGCAGCCAGTGGCTCTTGTAGAGAAAACT 718

Qy    826 TAGGATACCTCTCTCACTTTCTGTTTCTTGCCGTCCACCCCGGGCCATGCCAGTGTGTCC 885
Db    719 TAGGATACCTCTCTCACTTTCTGTTTCTTGCCGTCCACCCCGGGCCATGCCAGTGTGTCC 778

Qy    886 CTCTGGGTCCCTCCAAAACTCTGGTCAGTTCAAGGATGCCCCTCCCAGGCTATGCTTTTC 945
Db    779 TCTGGGTCCCCTCCAAAAATCTGGTCA-TTCAAGGATCCCCTCCCAAGGCTATGCTTTTC 837

Qy    946 TATAACTTTTAAATAAACCTTGGGGGTTGATGGAGTCAAAAA 987
Db    838 TATAACTTTTAAATAAACCTTGGGGGGTGAATGGAATAAAAA 879
```

```
RESULT   2
MMU86357
LOCUS       MMU86357      892 bp    mRNA            ROD       11-SEP-1997
DEFINITION  Mus musculus chemokine (TECK) mRNA, complete cds.
ACCESSION   U86357
NID         g2388628
VERSION     U86357.1  GI:2388628
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 892)
  AUTHORS   Vicari,A.P., Figueroa,D.J., Hedrick,J.A., Foster,J.S., Singh,K.P.,
            Menon,S., Copeland,N.G., Gilbert,D.J., Jenkins,N.A., Bacon,K.B. and
            Ziotnik,A.
  TITLE     TECK: a novel cc chemokine specifically expressed by thymic
            dendritic cells and potentially involved in T cell development
  JOURNAL   Immunology 7, 291-301 (1997)
REFERENCE   2  (bases 1 to 892)
  AUTHORS   Vicari,A.P. and Zlotnik,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JAN-1997) Immunology, DNAX Research Institute, 901
            California Ave., Palo Alto, CA 94304, USA
FEATURES             Location/Qualifiers
     source          1. .892
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="8"
                     /tissue_type="thymus"
     gene            1. .892
                     /gene="TECK"
     CDS             1. .435
                     /gene="TECK"
                     /codon_start=1
                     /product="chemokine"
                     /protein_id="AAB69982.1"
                     /db_xref="PID:g2388629"
                     /db_xref="GI:2388629"
                     /translation="MKLWLFACLVACFVGAWMPVVHAQGAFEDCCLGYQHRIKWNVLR
                     HARNYHQQEVSGSCNLRAVRFYFRQKVVCGNPEDMNVKRAIRILTARKRLVHWKSASD
                     SQTERKKSNHMKSKVENPNSTSVRSATLGHPRMVMMPRKTNN"
BASE COUNT      262 a    188 c    236 g    206 t
ORIGIN
```

Query Match 13.2%; Score 130.6; DB 12; Length 892;
Best Local Similarity 68.6%; Pred. No. 5.1e-23;
Matches 197; Conservative 0; Mismatches 84; Indels 6; Gaps 1;

```
Qy    106 ATGAACCTGTGGCTCCTGGCCTGCCTGGTGGCCGGCTTCCTGGGAGCCTGGGCCCCCGCT 165
Db      1 ATGAAACTGTGGCTTTTTGCCTGCCTGGTTGCCTGTTTTGTTGGGGCCTGGATGCCGGTT 60

Qy    166 GTCCACGCCCAAGGTGTCTTTGAGGACTGCTGCCTGGCCTACCACTACCCCATTGGGTGG 225
Db     61 GTCCATGCCCAAGGTGCCTTTGAAGACTGCTGCCTGGGTTACCAGCACAGGATCAAATGG 120

Qy    226 GCTGTGCTCCGGCGCGCCTGGACTTACCGGATCCAGGAGGTGAGCGGGAGCTGCAATCTG 285
```

```
           ||| ||||||| ||  ||| ||| |     ||||| ||||| || |||||||| ||
Db    121 AATGTTCTCCGGCATGCTAGGAATTATCACCAGCAGGAAGTGAGTGGAAGCTGCAACCTA 180

Qy    286 CCTGCTGCGATATTCTACCTCCCCAAGAGACACAGGAAGGTGTGTGGGAACCCCAAAAGC 345
          | ||||| || ||||||| ||| | ||| |          ||||||||||| ||  |   |
Db    181 CGTGCTGTGAGATTCTACTTCCGCCAGAAA------GTAGTGTGTGGGAATCCAGAGGAC 234

Qy    346 AGGGAGGTGCAGAGAGCCATGAAGCTCCTGGATGCTCGAAATAAGGT 392
          | | | ||| |||| || || |   || ||   ||| | || | | |
Db    235 ATGAATGTGAAGAGGGCGATAAGAATCTTGACAGCTAGGAAAAGGCT 281


RESULT   3
I66494
LOCUS       I66494       7218 bp    DNA             PAT       23-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
NID         g2724471
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN


  Query Match             5.2%;  Score 51.2;  DB 5;  Length 7218;
  Best Local Similarity   3.9%;  Pred. No. 0.0045;
  Matches   14;  Conservative  202;  Mismatches  140;  Indels    0;  Gaps    0;

Qy    601 TCTTTCTCCGATAAAACCGTCGCCCTACAGACCCAGCTGTCCCCACGCCTCTGTCTTTTG 660
          ::::::::: :    :: :: :::: :   :::  :: ::::: : ::::: ::::::
Db   1104 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1163

Qy    661 GGTCAAGTCTTAATCCCTGCACCTGAGTTGGTCCTCCCTCTGCACCCCCACCACCTCCTG 720
          ::   ::::  ::::: : :::   ::  :::::::::: : ::::: :: ::::::
Db   1164 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1223

Qy    721 CCCGTCTGGCAACTGGAAAGAGGGAGTTGGCCTGATTTTAAGCCTTTTGCCGCTCCGGGG 780
          ::: :::  :  ::            ::  :::  ::::   :::::: :: ::::
Db   1224 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1283

Qy    781 ACCAGCAGCAATCCTGGGCAGCCAGTGGCTCTTGTAGAGAAGACTTAGGATACCTCTCTC 840
          ::  :  :  ::::   :  ::  :  ::::: :        :::    : ::::::::
Db   1284 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1343

Qy    841 ACTTTCTGTTTCTTGCCGTCCACCCCGGGCCATGCCAGTGTGTCCCTCTGGGTCCCTCCA 900
          :::::: :::::: :: ::: ::::   :: : ::  : : :::::::   :::::::
Db   1344 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1403

Qy    901 AAACTCTGGTCAGTTCAAGGATGCCCCTCCCAGGCTATGCTTTTCTATAACTTTTA 956
            ::::  ::  :::     : :::::::   | |   | ||||||  |||| |
Db   1404 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCAAATTCTTCTATCTCTTTAA 1459


RESULT   4
AF046861/c
LOCUS       AF046861     2451 bp    DNA             MAM       02-OCT-1998
DEFINITION  Bos taurus somatotropin receptor gene, promoter, exon 1B and 1C.
ACCESSION   AF046861
NID         g3650493
VERSION     AF046861.1  GI:3650493
KEYWORDS    .
SOURCE      Bos taurus.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae;
            Bovinae; Bos.
REFERENCE   1  (bases 1 to 2451)
  AUTHORS   Lucy,M.C., Boyd,C.K. and Okamura,C.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-1998) Animal Science, University of Missouri, 164
            ASRC, Columbia, MO 65211, USA
REFERENCE   2  (bases 1 to 2451)
  AUTHORS   Lucy,M.C., Boyd,C.K. and Okamura,C.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-1998) Animal Science, University of Missouri, 164
            ASRC, Columbia, MO 65211, USA
  REMARK    Sequence update by submitter
COMMENT     On Sep 26, 1998 this sequence version replaced gi:2896014.
FEATURES             Location/Qualifiers
     source          1. .2451
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /chromosome="20"
     promoter        1. .1206
     CAAT_signal     1169. .1173
     mRNA            join(1207. .1703,1926. .>2451)
                     /product="somatotropin receptor"
     exon            1207. .1703
                     /note="1B; untranslated"
     exon            1926. .>2451
                     /note="1C; untranslated"
BASE COUNT      494 a    683 c    675 g    599 t
ORIGIN


  Query Match             4.7%;  Score 46.2;  DB 3;  Length 2451;
  Best Local Similarity   51.2%;  Pred. No. 0.082;
  Matches  108;  Conservative    0;  Mismatches  103;  Indels    0;  Gaps    0;

Qy     68 TTCGTCCAGGTGCCCAGGGAGGAGGACCCGCCTGCAGCATGAACCTGTGGCTCCTGGCCT 127
          | ||  | || |||||||| |  || |||  | ||  | ||    |   | | ||    |
Db   1665 TCCGCGCTGGGGCCCAGGGCGCGGGGCCCTGCGGCCCCCTGGGTCCCCGCCGCCGCATCG 1606

Qy    128 GCCTGGTGGCCGGCTTCCTGGGAGCCTGGGCCCCCGCTGTCCACGCCCAAGGTGTCTTTG 187
          ||   |  |||| |  |   |  |||   ||  | || |||  |||    |  | |   |
Db   1605 GCTCAGCAGCCGCCGCCGCCGCCGCCGCCGCTGCAGCAGTCTCCGCTGCCGCCGCCGCCG 1546

Qy    188 AGGACTGCTGCCTGGCCTACCACTACCCCATTGGGTGGGCTGTGCTCCGGCGCGCCTGGA 247
            |       || | ||  |  ||   ||  ||| || ||  |||||||  || |
Db   1545 CCGCTGCTCCCCAGACCGCCTCCTCTTCCCATGGCTGAGCCCTGCTCCGCGGCTCGCAAG 1486

Qy    248 CTTACCGGATCCAGGAGGTGAGCGGGAGCTG 278
          ||  | || ||  ||||| | | ||| || |
Db   1485 CTGTCAGGCTCGGGGAGGGGGGAGGGGGCGG 1455


RESULT   5
HSL261H12
LOCUS       HSL261H12   40198 bp    DNA             PRI       15-MAR-1996
DEFINITION  Human DNA sequence from cosmid L261H12, Huntington's Disease
            Region, chromosome 4p16.3 contains ESTs, STSs and a CpG island.
ACCESSION   Z54072
NID         g974828
VERSION     Z54072.1  GI:974828
KEYWORDS    4p16.3; CpG island.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 40198)
  AUTHORS   Buck,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-AUG-1995) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1RQ, England. E-mail: humquery@sanger.ac.uk
COMMENT     IMPORTANT:  This sequence is not the entire insert of clone
            L261H12.  It may be shorter because we only sequence overlapping
```

```
          sections once, or longer because we arrange for a small overlap
          between neighbouring submissions.
          The true left end of clone L261H12 is at 1.
          The true right end of clone L261H12 is at 40102.
          The true right end of clone L247F6 is at 16300.
          The true left end of clone  L145E5 is at 25120.
          The end of this sequence (40099. .40198) overlaps with the start of
          sequence HSL145E5.
          L261H12 is from cosmid library LA04NC01 constructed at the Human
          Genome Center, Los Alamos National Laboratory, NM 87545 under the
          auspices of the U.S. Department of Energy. The library was
          constructed using flow-sorted human chromosome 4 from a
          Hamster-Human hybrid cell line ( UV20HL21-27 ) containing human
          chromosomes 4, 8 and 21.
          VECTOR: sCos1
          L261H12 is contained in a clone contig spanning
          2Mb which is described in Baxendale et al, Nature Genetics 4 ( 1993
          ) 181- 186.  See also Myers et al, Cytogenet Cell Genet. 66 ( 1994
          ) 218-230.
FEATURES             Location/Qualifiers
     source          1. .40198
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4p16.3"
                     /clone="L261H12"
                     /clone_lib="LA04NC01"
                     /cell_line="UV20HL21-27"
     repeat_region   1494. .1523
                     /note="15 copies of 2 mer 93.3 % conserved"
     repeat_region   6624. .7015
                     /note="8 copies of 49 mer 84.9 % conserved"
     repeat_region   7275. .7338
                     /note="MLT1B element fragment"
     repeat_region   7541. .7681
                     /note="MLT1 element fragment"
     repeat_region   9098. .9157
                     /note="3 copies of 20 mer 85.0 % conserved"
     repeat_region   13005. .13328
                     /note="Low complexity region: TA, CA rich"
     repeat_region   13331. .13616
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   13618. .13906
                     /note="Alu repeat: matches 308. .1 of consensus"
     misc_feature    complement(14195. .14419)
                     /note="matches EST T19703"
     misc_feature    complement(14229. .14366)
                     /note="matches EST T19718"
     repeat_region   20250. .20317
                     /partial
                     /note="Alu repeat: matches 1. .77 of consensus"
     misc_feature    21991. .22963
                     /note="putative CpG island"
     unsure          22312. .22885
                     /note="78% GC - base count and sequence not completely
                     certain"
     repeat_region   24933. .25052
                     /note="SVA element fragment"
     repeat_region   25252. .25437
                     /note="SVA element fragment"
     repeat_region   26386. .26667
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   29548. .29752
                     /partial
                     /note="Alu repeat: matches 86. .308 of consensus"
     repeat_region   30075. .30178
                     /note="Inverted repeat"
     repeat_region   30657. .30945
                     /note="Alu repeat: matches 1. .308 of consensus"
     repeat_region   30978. .31127
                     /note="MER25 element fragment"
     repeat_region   31168. .31230
                     /note="MER25 element fragment"
     repeat_region   32036. .32139
                     /note="L1 element fragment"
     repeat_region   32197. .32611
                     /note="L1 element fragment"
     repeat_region   32730. .33323
                     /note="L1 element fragment"
     repeat_region   33322. .33783
                     /note="MER4A2 element fragment"
     misc_feature    34275. .34637
                     /note="matches STS L09772 - contains CA repeat and Alu
                     fragment"
     repeat_region   34276. .34359
                     /note="Low complexity region: T rich.  A, C poor."
     misc_feature    34392. .34661
                     /note="matches STS L03349 - contains CA repeat and Alu
                     fragment"
     repeat_region   34454. .34491
                     /note="19 copies of CA 100.0 % conserved"
     repeat_region   34579. .34818
                     /partial
                     /note="Alu repeat: matches 255. .1 of consensus"
     repeat_region   35147. .35192
                     /note="23 copies of 2 mer 80.4 % conserved"
     repeat_region   35257. .35547
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   35635. .35927
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   36020. .36223
                     /partial
                     /note="Alu repeat: matches 1. .222 of consensus"
     repeat_region   36422. .36714
                     /note="Alu repeat: matches 1. .308 of consensus"
     repeat_region   36860. .36999
                     /note="LTR9 element fragment"
     repeat_region   37055. .37096
                     /note="LTR9 element fragment"
     repeat_region   37709. .38000
                     /note="Alu repeat: matches 1. .308 of consensus"
     repeat_region   38345. .38634
                     /note="Alu repeat: matches 1. .308 of consensus"
     repeat_region   38860. .39150
                     /note="Alu repeat: matches 1. .308 of consensus"
     repeat_region   39193. .39470
                     /note="Alu repeat: matches 1. .308 of consensus"
     repeat_region   39531. .39822
                     /note="Alu repeat: matches 308. .1 of consensus"
BASE COUNT     9161 a  11896 c  11011 g   8130 t
ORIGIN


  Query Match             4.3%;  Score 42.8;  DB 10;  Length 40198;
  Best Local Similarity   50.0%;  Pred. No. 0.58;
  Matches  107;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy     688 TTGGTCCTCCCTCTGCACCCCCACCACCTCCTGCCCGTCTGGCAACTGGAAAGAGGGAGT 747
           | || ||  || | |||  |   || || ||| || ||| | |   ||  | | ||| ||
Db    3234 TGGGCCCCACCCCAGCAGGCTGGCCCCCGCCTCCCTGTCGGCCGCATGTGATGGGGGGGT 3293

Qy     748 TGGCCTGATTTTAAGCCTTTTGCCGCTCCGGGGACCAGCAGCAATCCTGGGCAGCCAGTG 807
           ||   | || ||||  |       |  |   | ||  |  |  | | | |||| | |
Db    3294 GGGGGGGGTTCTAAGTATCACATTCCCTCTCTGGCCTCCCCCTGTGCCTGCCAGCTCGGG 3353

Qy     808 GCTCTTGTAGAGAAGACTTAGGATACCTCTCTCACTTTCTGTTTCTTGCCGTCCACCCCG 867
           |  ||   |  | ||||  |||  ||    ||||    ||||| | |   |    | | |
Db    3354 GGCCTCTCAAGGCAGACAGAGGCCACTCTCCTCAGAGCCTGTTGCCTTTTGCAGGCTCAG 3413

Qy     868 GGCCATGCCAGTGTGTCCCTCTGGGTCCCTCCAA 901
           |||    || ||| |  ||   ||    | || |
Db    3414 GGCACCTCCTGTGGGCTCCCACGGCCAGCCCCGA 3447


RESULT   6
STMDNAGYR/c
```

```
LOCUS       STMDNAGYR    8200 bp    DNA             BCT       20-APR-1995
DEFINITION  Streptomyces coelicolor dnaA-gyrA region.
ACCESSION   L27063
NID         g436023
VERSION     L27063.1  GI:436023
KEYWORDS    6-phosphogluconate dehydrogenase homolog; DNA gyrase; dnaA gene;
            gyrA gene; gyrB gene.
SOURCE      Streptomyces coelicolor (strain A3(2)) DNA.
  ORGANISM  Streptomyces coelicolor
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 8200)
  AUTHORS   Calcutt,M.J.
  TITLE     Gene organization in the dnaA-gyrA region of the Streptomyces
            coelicolor chromosome
  JOURNAL   Gene 151 (1-2), 23-28 (1994)
  MEDLINE   95129865
FEATURES             Location/Qualifiers
     source          1. .8200
                     /organism="Streptomyces coelicolor"
                     /strain="A3(2)"
                     /db_xref="taxon:1902"
     gene            1. .36
                     /gene="dnaN"
     CDS             <1. .36
                     /gene="dnaN"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA65211.1"
                     /db_xref="PID:g777777"
                     /db_xref="GI:777777"
                     /translation="KYLIMPVRLSG"
     repeat_region   21. .60
                     /note="putative"
                     /rpt_type=tandem
     CDS             280. .1155
                     /note="product homolog; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="6-phosphogluconate dehydrogenase"
                     /protein_id="AAA65212.1"
                     /db_xref="PID:g436024"
                     /db_xref="GI:436024"
                     /translation="MELGLVGLGKMGGNMRERIRRAGHTVVGYDRNPDLADVHSLEEL
                     VGKLSAPRVVWVMVPAGEATQATVDRLGELLEPGDVVVDGGNSRWTDDEKHAEELAAK
                     GIGFVDCGVSGGVWGLENGYALMYGGDAENVAKVQPVFDALKPEGDFGAVHAGKVGAG
                     HFAKMVHNGIEYAMMQAYAEGWELLEKVDSVENVREVFRSWQEGTVIRSWLLDLAVNA
                     LDDDEHLDGLRGYAQDSGEGRWTVEAAIDNAVPLPAITASLFARFASRQDDSPQMKMI
                     AALRNQFGGHAVEKK"
     gene            1186. .2307
                     /gene="recF"
     CDS             1186. .2307
                     /gene="recF"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA65213.1"
                     /db_xref="PID:g436025"
                     /db_xref="GI:436025"
                     /translation="MHVTHLSLADFRSYARVEVPLDPGVTAFVGPNGQGKTNLVEAVG
                     YLATLGSHRVSSDAPLVRMGAERAVIRAQVRQGERQQLIELELNPGRANRARVNRSSQ
                     VKPRDVLGIVRTVLFAPEDLALVKGDPGERRRFLDELITARSPRMAGVRSDYDRVLKQ
                     RNTLLKSAALARRHGGRTMDLSTLDVWDQHLARAGAELLAQRLDLIASVQPLADKAYE
                     QLAPGGGPVALEYKPSAPGEAHTREDLYEQLMAALAEARKQEIERGVTLVGPHRDDLL
                     LKLGSLPAKGYASHGESWSYALALRLASFDLLRAEGNEPVLVLDDVFAELDARRRERL
                     AELVAPGEQVLVTAAVDDDVPHVLAGARFTVAEGTVERV"
     CDS             2304. .2876
                     /note="ORF191; putative"
                     /codon_start=1
                     /transl_table=11
                     /function="unknown"
                     /protein_id="AAA65214.1"
                     /db_xref="PID:g436026"
                     /db_xref="GI:436026"
                     /translation="MSADAPEPGPGQSPGERASGGPEPSGVDLARVALRAAREAARAR
                     GDAAQQKKQARRGGLRSGARADGRDPMALGSAINRLITERGWETPAAVGGVMGRWPEI
                     VGADVAKHCVPERYDEDERVLVVRCDSTAWATNLRLLAPTLVARLNEDLGHGSVRMIK
                     VLGPGGPGGPGRRYGPLRAPGSQGPGDTYG"
     gene            3311. .5371
                     /gene="gyrB"
     CDS             3311. .5371
                     /gene="gyrB"
                     /note="putative"
                     /codon_start=1
                     /transl_except=(pos:3311. .3313,aa:Val)
                     /transl_table=11
                     /product="DNA gyrase"
                     /protein_id="AAA65215.1"
                     /db_xref="PID:g436027"
                     /db_xref="GI:436027"
                     /translation="VADSGNPNENNPSTDTGVNDAVSTSHGDASASYDASAITVLEGL
                     DAVRKRPGMYIGSTGERGLHHLVQEVVDNSVDEALAGHADTIDVTILPDGGVRVVDNG
                     RGIPVGIVPSEGKPAVEVVLTVLHAGGKFGGGGYAVSGGLHGVGVSVVNALSTRVAVE
                     VKTDGYRWTQEYKLGVPTASLARHEATEETGTTVTFWADGDIFETTDYSFETLSRRFQ
                     EMAFLNKGLKINLTDERESAKATAGADEAGEDEKHEVKSVSYHYEGGIVDFVTYLNSR
                     KGELVHPTVIDLEAEDKDKSLSLEVAMQWNGGYTEGVYSFANIIHTHEGGTHEEGFRA
                     ALTSLINKYARDKKLLREKDDNLTGDDIREGLTAIISVKLAEPQFEGQTKTKLGNTEV
                     KTFVQKVVYEHLTDWLDRNPNEAADIIRKGIQAAHARVAARKARDLTRRKGLLESASL
                     PGKLSDCQSNDPTKCEIFIVEGDSAGGSAKSGRNPQYQAILPIRGKILNVEKARIDRI
                     LQNQEIQAMISAFGTGVHEDFDIEKLRYHKIILMADADVDGQHINTLLLTFLFRFMRP
                     LVESGHVYLSRPPLYKIKWGRDDFEYAYSDRERDALIEMGRQAGKRIREDSVQRFKGL
                     GEMNAEELRITTMDQEHRVLGQVTLDDAAQADDLFSVLMGEDVEARRAFIQRNAKDVR
                     FLDI"
     gene            5414. .8008
                     /gene="gyrA"
     CDS             5414. .8008
                     /gene="gyrA"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="DNA gyrase"
                     /protein_id="AAA65216.1"
                     /db_xref="PID:g436028"
                     /db_xref="GI:436028"
                     /translation="MTDENTPVTTPEGDALAMRVEPVGLETEMQRSYLDYAMSVIVSR
                     ALPVVRDGLKPVHRRVLYAMYDGGYRPERGFYKCARVVGDVMGNYHPHGDSSIYDALV
                     RLAQPWSMRMPLVDSNGNFGSPGNDPAAAMRYTECKMAPLSMEMVRDIDEETVDFTDN
                     YDGRSQEPTVLPARFPNLLINGSAGIAVGMATNIPPHNLREVAAGAQWYLENYEASHE
                     ELLDALIERIKGPDFPTGALVVGRKGIEEAYRTGRGSITMRAVVEVEEIQNRQCLVVT
                     ELPYQTNPDNLAQKIADLVKDGKVGGIADVRDETSSRTGQRLVIVLKRDAVAKVVLNN
                     LYKHTDLQSNFGANMLALVDGVPRTLSLDAFIRHWVNHQIEVIVRRTRFRLRKAEERA
                     HILRGLLKALDAIDEVIALIRRSDTVEIARGGLMDLLEIDEIQANAILEMQLRRLAAL
                     ERQKIVREHDELQAKITEYNEILASPVRQRGIVSEELTALVEKYGDDRKTKLIPYEGD
                     MSIEDLIAEEDIVVTVTRGGYIKRTKTDDYRAQKRGGKGVRGTKLKEDDIVNHFFVST
                     THHWLLFFTNKGRVYRAKAYELPDAGRDARGQHVANLLAFQPDETIAQIRAIRDYEAV
                     PYLVLATKAGLVKKTPLKDYDSPRSGGVIAINLREQADGSDDELIGAELVSAEDDLLL
                     ISKKAQSIRFTASDDTLRPMGRATSGVKGMSFREGDELLSMNVVRAGTFVFTATDGGY
                     AKRTSVDEYRVQGRGGLGIKAAKIVEDRGSLVGALVVEEHDEILAITLSGGVIRTRVN
                     GVRETGRDTMGVQLINLGKRDAVVGIARNAEAGREAEEVDGDVAVDETAEGAATTGTD
                     EGEAPSAE"
BASE COUNT     1410 a   2773 c   2910 g   1107 t
ORIGIN


  Query Match             4.1%;  Score 41;  DB 1;  Length 8200;
  Best Local Similarity   47.5%;  Pred. No. 1.6;
  Matches  122;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy      15 CAGCCCGGCGGGCATCAGCTCCCTTGACCCAGTGGATATCGGTGGCCCCGTTATTCGTCC 74
           ||   | || | |  | | | || |  || || || || |||  |  |||     |||
Db    8096 CACGGCCGCCGCCGCCCGTTTCCGTACCCGAGGGGGTACCGGCCGATCCGGCGCCCGTGG 8037

Qy      75 AGGTGCCCAGGGAGGAGGACCCGCCTGCAGCATGAACCTGTGGCTCCTGGCCTGCCTGGT 134
             || || |  |     | |  |        |   ||   ||| ||| |||  | |
Db    8036 CTCCGCTCACGATGACTCTCTCCTCGTGCTACTCGGCCGAGGGCGCCTCGCCCTCGTCCG 7977
```

```
Qy     135 GGCCGGCTTCCTGGGAGCCTGGGCCCCCGCTGTCCACGCCCAAGGTGTCTTTGAGGACTG 194
           |||||     |  ||  ||   | |       ||| ||| |||    | | | ||   |
Db    7976 TGCCGGTGGTCGCGGCACCCTCGGCGGTCTCGTCGACGGCCACATCGCCGTCGACCTCCT 7917

Qy     195 CTGCCTGGCCTACCACTACCCCATTGGGTGGGCTGTGCTCCGGCGCGCCTGGACTTACCG 254
           | |||| || | || |  |  | ||| ||| | ||    |    ||  |  |  |  ||
Db    7916 CCGCCTCGCGTCCCGCCTCGGCGTTGCGTGCGATGCCGACGACGGCATCGCGCTTGCCCA 7857

Qy     255 GATCCAGGAGGTGAGCG 271
           |||  |  || ||| ||
Db    7856 GATTGATCAGTTGAACG 7840


RESULT   7
HSAJ3461
LOCUS       HSAJ3461      345 bp    mRNA            PRI       20-JAN-1999
DEFINITION  Homo sapiens mRNA for MPIF-2 gene variant, partial.
ACCESSION   AJ223461
NID         g4165481
VERSION     AJ223461.1  GI:4165481
KEYWORDS    eotaxin; MPIF-2 gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 345)
  AUTHORS   Hein,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-1998) Hein H., Dermatology/Hautkilnik,
            Christian-Albrechts-Universitaet zu Kiel, Schittenhelmstr. 7, Kiel,
            D-24105, FRG
REFERENCE   2  (bases 1 to 345)
  AUTHORS   Hein,H. and Theran,L.
  TITLE     cDNA,genomic organisation and chromosomal location of the MPIF-2
            (eotaxin-2) gene
  JOURNAL   Unpublished
COMMENT     Related sequene: U85768.
FEATURES             Location/Qualifiers
     source          1. .345
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="monocyte"
     gene            1. .345
                     /gene="MPIF-2"
     CDS             <1. .>345
                     /gene="MPIF-2"
                     /codon_start=1
                     /protein_id="CAA11383.1"
                     /db_xref="PID:e1371624"
                     /db_xref="PID:g4165482"
                     /db_xref="GI:4165482"
                     /translation="GLMTIVTSLLFLGVCAHHIIPTGSVVIPSPCCMFFVSKRIPENR
                     VVSYQLSSRSTCLKAGVIFTTKKGQQFCGDPKQEWVQRYMKNLDAKQKKASPRARAVA
                     VKGPVQRYPGNQT"
BASE COUNT       79 a    104 c     92 g     70 t
ORIGIN


  Query Match             4.1%;  Score 40.2;  DB 9;  Length 345;
  Best Local Similarity   56.4%;  Pred. No. 2.7;
  Matches   75;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy     258 CCAGGAGGTGAGCGGGAGCTGCAATCTGCCTGCTGCGATATTCTACCTCCCCAAGAGACA 317
           ||||  |   ||| |||||      ||     || |   |  ||| |  | ||||||
Db     144 CCAGCTGTCCAGCAGGAGCACATGCCTCAAGGCAGGAGTGATCTTCACCACCAAGAAGGG 203

Qy     318 CAGGAAGGTGTGTGGGAACCCCAAAAGCAGGGAGGTGCAGAGAGCCATGAAGCTCCTGGA 377
           |  | || | |||||  |||||||      |  ||| |||||   |||||||  ||||||
Db     204 CCAGCAGTTCTGTGGCGACCCCAAGCAGGAGTGGGTCCAGAGGTACATGAAGAACCTGGA 263

Qy     378 TGCTCGAAATAAG 390
           ||     | |||
Db     264 CGCCAAGCAGAAG 276


RESULT   8
MTCY22G8
LOCUS       MTCY22G8    22550 bp    DNA             BCT       17-JUN-1998
DEFINITION  Mycobacterium tuberculosis H37Rv complete genome; segment 49/162.
ACCESSION   Z95585 AL123456
NID         g3261787
VERSION     Z95585.1  GI:3261787
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium.
REFERENCE   1  (bases 1 to 22550)
  AUTHORS   Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
            Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
            Tekaia,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
            Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S.,
            Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
            Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
            Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
            Squares,S., Sqares,R., Sulston,J.E., Taylor,K., Whitehead,S. and
            Barrell,B.G.
  TITLE     Deciphering the biology of Mycobacterium tuberculosis from the
            complete genome sequence
  JOURNAL   Nature 393 (6685), 537-544 (1998)
  MEDLINE   98295987
  REMARK    Erratum:[[published erratum appears in Nature 1998 Nov
            12;396(6707):190]]
REFERENCE   2  (bases 1 to 22550)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
            tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
            Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
            Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
            75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
COMMENT     On Jun 27, 1998 this sequence version replaced gi:2117205.
            Notes:
            Details of M. tuberculosis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
            been renumbered from the original cosmid submissions but the old
            gene designations are in brackets after the new gene numbers.
            Gene prediction was based on a Hidden Markov Model of TB genes
            implemented in TBparse (Krogh) supplemented with visual inspection
            of positional base preference in codons, especially where there is
            an increase in the observed/expected third position G + C.
            CAUTION:  In some cases we may not have predicted the correct
            initiation codon.  Where possible we choose an initiation codon
            (atg, gtg, or ttg) which is preceded by an upstream ribosome
            binding site sequence (optimally 5-13bp before the initiation
            codon).  If this cannot be identified we choose the most upstream
            initiation codon.
FEATURES             Location/Qualifiers
     source          1. .22550
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37Rv"
                     /db_xref="taxon:1773"
                     /clone="Y22G8"
     RBS             53. .56
                     /note="possible RBS upstream of Rv1113"
     gene            64. .261
                     /gene="Rv1113"
     CDS             64. .261
                     /gene="Rv1113"
                     /note="Rv1113, (MTCY22G8.02), len: 65. Unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein Rv1113"
                     /protein_id="CAB09045.1"
```

```
                /db_xref="PID:e317138"
                /db_xref="PID:g2117207"
                /db_xref="GI:2117207"
                /db_xref="SPTREMBL:O06565"
                /translation="MRTTVTVDDALLAKAAELTGVKEKSTLLREGLQTLVRVESARRL
                AALGGTDPQATAAPRRRTSPR"
gene            258. .632
                /gene="Rv1114"
CDS             258. .632
                /gene="Rv1114"
                /note="Rv1114, (MTCY22G8.03), len: 124. Slight
                similarityto M.tuberculosis hypothetical gene MTCY159.08c
                (33.0% identity in 115 aa overlap)"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv1114"
                /protein_id="CAB09046.1"
                /db_xref="PID:e317242"
                /db_xref="PID:g2117208"
                /db_xref="GI:2117208"
                /db_xref="SPTREMBL:O06566"
                /translation="MILVDTSVWIEHLRAADARLVELLGDDEAGCHPLVIEELALGSI
                KQRDVVLDLLANLYQFPVVTHDEVLRLVGRRRLWGRGLGAVDANLLGSVALVGGARLW
                TRDKRLKAACAESGVALAEEVS"
RBS             826. .828
                /note="possible RBS upstream of Rv1115"
gene            835. .1533
                /gene="Rv1115"
CDS             835. .1533
                /gene="Rv1115"
                /note="Rv1115, (MTCY22G8.04), len: 232. Unknown"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv1115"
                /protein_id="CAB09047.1"
                /db_xref="PID:e317139"
                /db_xref="PID:g2117209"
                /db_xref="GI:2117209"
                /db_xref="SPTREMBL:O06567"
                /translation="MISTTRIDFLWILSVAFASMIALATLLTLINQVVGTPYIPGGDS
                PAGTDCSELASWVSNAATARPVFGDRFNTGNEEAALAARGFQQGTAPNALVIGWNGHH
                TAVTLPDGTPVSSGEGGGVRVGGGGAYQPKFTHHMYLPMDVDAGEDQPPAPDEPVTAV
                DDVEPEMPAPCPTQRPPVTPRHNLCNKLRTMPGALSAALAAAAPVWPAPISGCRGFST
                SLLAKRNHPVIVGK"
gene            1651. .1836
                /gene="Rv1116"
CDS             1651. .1836
                /gene="Rv1116"
                /note="Rv1116, (MTCY22G8.05), len: 61. Unknown"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv1116"
                /protein_id="CAB09048.1"
                /db_xref="PID:e317140"
                /db_xref="PID:g2117210"
                /db_xref="GI:2117210"
                /db_xref="SPTREMBL:O06568"
                /translation="MCSRMADEPRLEAGAHPFEEGRDKAPELRATQMDHVRFTEGRRE
                RNRDRLERSQQFRQPGR"
RBS             2273. .2275
                /note="possible RBS upstream of Rv1117"
gene            2281. .2604
                /gene="Rv1117"
CDS             2281. .2604
                /gene="Rv1117"
                /note="Rv1117, (MTCY22G8.06), len: 107. Some similarity
                toB. subtilis hypothetical protein TR:P94425 (EMBL:D50453)
                YCNE (95 aa), fasta scores; opt: 128 z-score: 237.3 E():
                5.1e-06, 28.3% identity in 92 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv1117"
                /protein_id="CAB09049.1"
                /db_xref="PID:e317243"
                /db_xref="PID:g2117211"
                /db_xref="GI:2117211"
                /db_xref="SPTREMBL:O06569"
                /translation="MIFIVVKFETKPEWTERWPDLVASFTAATRAEEGNLWFEWSRSL
                DDPAEYVLVESFRDGEAGGVHVNSDHFRQAMRELPKALASTPKIISQTIDATGWSAMG
                EMTVG"
gene            complement(2619. .3479)
                /gene="Rv1118c"
CDS             complement(2619. .3479)
                /gene="Rv1118c"
                /note="Rv1118c, (MTCY22G8.07c), len: 286. Unknown"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv1118c"
                /protein_id="CAB09050.1"
                /db_xref="PID:e317141"
                /db_xref="PID:g2117212"
                /db_xref="GI:2117212"
                /db_xref="SPTREMBL:O06570"
                /translation="MQSGPHLVGRVGTSFPLIARHQGATRDDAGDTGQPDPLPHVAHP
                DRLYPPMVHGVDPSTLALDRALNETRTGDLWLFRGRSRPDRAIQTLTNAPVNHVGMTV
                AIDDLPPLIWHAELGDKLLDVWTGTNHRGVQLNDARQVVQQWAGRYRQRCWLRQLTPH
                ANRDQEDKLLRVIARMNGTPFPTTARLTGRWLRGRLPTLNDWLRGIPVLDRKVREQTQ
                RRKQQQRTMGLATAYCAETVAITYEEMGLLVTDKDAHWFDPGKFWSGDSLPLAPGYRL
                GHEIAVDVGG"
gene            complement(3512. .3661)
                /gene="Rv1119c"
CDS             complement(3512. .3661)
                /gene="Rv1119c"
                /note="Rv1119c, (MTCY22G8.08c), len: 49. Unknown"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv1119c"
                /protein_id="CAB09051.1"
                /db_xref="PID:e317142"
                /db_xref="PID:g2117213"
                /db_xref="GI:2117213"
                /db_xref="SPTREMBL:O06571"
                /translation="MTARVAGQAVGGQILVGEPVHDAVSDCADIRFGSYRLFSLDAAP
                GPDLD"
gene            complement(3658. .4152)
                /gene="Rv1120c"
CDS             complement(3658. .4152)
                /gene="Rv1120c"
                /note="Rv1120c, (MTCY22G8.09c), len: 164. Some
                similarityto YZ19_MYCTU Q10551 hypothetical 31.2 kd
                protein cy31.19c(285 aa), fasta scores; opt:  156 z-score:
                195.2 E(): 0.00074, 32.7% identity in 107 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv1120c"
                /protein_id="CAB09052.1"
                /db_xref="PID:e317244"
                /db_xref="PID:g2117214"
                /db_xref="GI:2117214"
                /db_xref="SPTREMBL:O06572"
                /translation="MLSGGREAVKTVWQTANLVRKEGFGAAVRSSIEDPADWAEVERP
                DLARVTPDGRVVILFSDIEESTALDERIGDRTWVKLIGAHDKLVHELVRRWSGHMVTS
                QGDGFMIAFARAEQAVRCGIDIQDALRNSAKRKRNQGIRVRIGTTWGARCGTVTICSA
                ATSQ"
RBS             4343. .4347
                /note="possible RBS upstream of Rv1121"
gene            4355. .5755
                /gene="zwf"
CDS             4355. .5755
                /gene="zwf"
                /note="Rv1121, (MTCY22G8.10), zwf, len 466. Probable
                glucose-6-phosphate 1-dehydrogenase, highly similar to eg.
                G6PD_ECOLI P22992 glucose-6-phosphate 1-dehydrogenase (EC
                1.1.1.49) (491 aa), fasta scores; opt:  642 z-score: 767.5
                E(): 0, 35.8% identityin 478 aa overlap"
                /codon_start=1
```

```
                /transl_table=11
                /product="zwf"
                /protein_id="CAB09053.1"
                /db_xref="PID:e317143"
                /db_xref="PID:g2117215"
                /db_xref="GI:2117215"
                /db_xref="SPTREMBL:O06573"
                /translation="MVDGGGGASDLLVIFGITGDLARKMTFRALYRLERHQLLDCPIL
                GVASDDMSVGQLVKWARESIGRTEKIDDAVFDRLAGRLSYLHGDVTDSQLYDSLAELI
                GSACRPLYYLEMPPALFAPIVENLANVRLLERARVAVEKPFGHDLASALELNARLRAV
                LGEDQILRVDHFLGKQPVVELEYLRFANQALAELWDRNSISEIHITMAEDFGVEDRGK
                FYDAVGALRDVVQNHLLQVLALVTMEPPVGSSADDLNDKKAEVFRAMAPLDPDRCVRG
                QYLGYTEVAGVASDSATETYVALRTEIDNWRWAGVPIFVRAGKELPAKVTEVRLFLRR
                VPALAFLPNRRPAEPNQIVLRIDPDPGMRLQISAHTDDSWRDIHLDSSFAVDLGEPIR
                PYERLLYAGLVGDHQLFAREDSIEQTWRIVQPLLDNPGEIHRYDRGSWGPEAAQSLLR
                GHRGWQSPWLPRGTDA"
RBS             5762. .5765
                /note="possible RBS upstream of Rv1122"
gene            5777. .6799
                /gene="gnd2"
CDS             5777. .6799
                /gene="gnd2"
                /note="Rv1122, (MTCY22G8.11), gnd2, len: 340. Probable
                6-phosphogluconate dehydrogenase, decarboxylating, highly
                similar to eg. Q53917 6-PHOSPHOGLUCONATE DEHYDROGENASE
                (291 aa)from Streptomyces coelicolor, fasta scores; opt:
                431 z-score: 492.0 E(): 2.2e-20, 44.5% identity in 335 aa
                overlap.Also similar to M. tuberculosis gnd MTCY359.29
                (33.0% identity in 351 aa overlap). Note that MTCY359.29
                (gnd) is mostsimilarto gnd's from Gram negative organisms,
                while gnd2 is most similar tognd's from Gram positive
                organisms"
                /codon_start=1
                /transl_table=11
                /product="gnd2"
                /protein_id="CAB09054.1"
                /db_xref="PID:e317144"
                /db_xref="PID:g2117216"
                /db_xref="GI:2117216"
                /db_xref="SPTREMBL:O06574"
                /translation="MQLGMIGLGRMGANIVRRLAKGGHDCVVYDHDPDAVKAMAGEDR
                TTGVASLRELSQRLSAPRVVWVMVPAGNITTAVIEELANTLEAGDIVIDGGNTYYRDD
                LRHEKLLFKKGIHLLDCGTSGGVWGRERGYCLMIGGDGDAFARAEPIFATVAPGVAAA
                PRTPGRDGEVAPSEQGYLHCGPCGSGHFVKMVHNGIEYGMMASLAEGLNILRNADVGT
                RVQHGDAETAPLPNPECYQYDFDIPEVAEVWRRGSVIGSWLLDLTAIALRESPDLAEF
                SGRVSDSGEGRWTAIAAIDEGVPAPVLTTALQSRFASRDLDDFANKALSAMRKQFGGH
                AEKPAN"
gene            complement(6792. .7700)
                /gene="bpoB"
CDS             complement(6792. .7700)
                /gene="bpoB"
                /note="Rv1123c, (MTCY22G8.12c), len: 302. bpoB, Some
                similarity to a range of enzymes including; DEH1_MORSP
                Q01398 haloacetatedehalogenase (294 aa), fasta scores; 64
                opt:  201z-score: 241.0 E(): 2.1e-06, 35.8% identity in
                134 aa overlap, and; BPA1_STRAU P33912 non-haem
                bromoperoxidase bpo-a1(274aa), fasta scores; opt:  187
                z-score: 225.2 E(): 1.6e-05,23.1% identity in 281 aa
                overlap, contains PS00216 Sugartransport proteins
                signature 1. Similar to several other M.tuberculosis
                proteins eg. MTCY13E12.26c (39.5% identity
                in256aaoverlap)"
                /codon_start=1
                /transl_table=11
                /product="bpoB"
                /protein_id="CAB09055.1"
                /db_xref="PID:e317245"
                /db_xref="PID:g2117217"
                /db_xref="GI:2117217"
                /db_xref="SPTREMBL:O06575"
                /translation="MTIWRVPSKVTSGPVSAVSSSPQAVAFSGARGITLVADEWNRGA
                AAADRPTILMLHGGGQNRFSWKNTGQILADEGHHVVALDTRGPGDSDRAPGADYAVET
                PTTDVLHVVEAIGRRVVVVEASMGGLTGILVAERAGPQTVNGLVLVDVVPRYEKEGNA
                RIRDFMLGNIDGFGSLEEAADAVAEYLPHRDKPRSPEGLKRNLRLRDGRWHWHWDPAM
                MTAPGHDPQLRTENFERAAMGLTIPVLLIRGKLSDVVSSDGARDFLAKVPNAEFVELS
                NAGRTAAGDDNDAFTDVVVDFVRRLS"
misc_feature    complement(7329. .7382)
                /gene="bpoB"
                /note="PS00216 Sugar transport proteins signature 1"
RBS             complement(7709. .7717)
                /note="possible RBS upstream of Rv1123c"
RBS             7770. .7772
                /note="possible RBS upstream of Rv1124"
gene            7775. .8725
                /gene="ephC"
CDS             7775. .8725
                /gene="ephC"
                /note="Rv1124, (MTCY22G8.13), len: 316. ephC, Similar to
                Q42566epoxide hydrolase (321 aa) from Arabidopsis
                thaliana, fastascores; opt:  298 z-score: 356.0 E():
                8.2e-13, 27.6% identity in 333 aa overlap. Similar to
                other M. tuberculosishydrolaseseg. MTCY09F9.26c (26.7%
                identity in 333 aa overlap)"
                /codon_start=1
                /transl_table=11
                /product="ephC"
                /protein_id="CAB09056.1"
                /db_xref="PID:e317145"
                /db_xref="PID:g2117218"
                /db_xref="GI:2117218"
                /db_xref="SPTREMBL:O06576"
                /translation="MRAGRGERESTWRTTMAEPHWIDVKGPNGDLKALTWGPAGAPVA
                LCLHGFPDTAYGWRKVAPRLAESGWHVVAPFMRGYAPSSIPADGSYHVGALMHDALRV
                RSAAGGTERDVIIGHDWGAIAATGLAAMPDSPFAKAVIMSVPPSAAFRPLGRVPERGR
                LLRELPHQLLRSWYILYFQLPWLPERSASWVVPLLWRRWSPGYHAEEDLRHVDAAIGT
                PEGRRAALGPYRATMRNTRAPADYADLNRLWTEAPKLPVLYLHGHDDGCATSAFTHWT
                ARVLPAGSEVAVVEHAGHFLQLEQPDKIAELIVAFIGSPG"
RBS             8725. .8727
                /note="possible RBS upstream of Rv1125"
gene            8730. .9974
                /gene="Rv1125"
CDS             8730. .9974
                /gene="Rv1125"
                /note="Rv1125, (MTCY22G8.14), len: 414.unknown"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv1125"
                /protein_id="CAB09057.1"
                /db_xref="PID:e317146"
                /db_xref="PID:g2117219"
                /db_xref="GI:2117219"
                /db_xref="SPTREMBL:O06577"
                /translation="MAGHRMAAVDAQFYWMSAKVPNDQFLLYAFDGEPTDLERAVAQV
                YRRARGCPGLGMRVQDRGALAYPQWVPTPVQRDQLVCHDLADRSWQGCLAAVVGLASK
                QLDMRRMPWRLHVFTPVHDVPGVSGLGTVAVMQFAHALGDGARASAMAAWLFGRPAAV
                PEIARSRAGFLPWRAAHAARAHLRLVRDTNAGLVAPGVGSRPPLSTNARPEGVRAVRT
                LLRRRSQLAGPTVTVTVLAAVSTGLLGLLGGDVDTLGAEVPMAKPGVPRSYNHFGNVV
                VGLYPRLEPDERVRRIATDLANARRRFEHPAMLSADRAFAAVPAALLRWGVSQFDAEV
                RPVRVAGNTVVSSVYRGAADLSFGDAPVVLTAGYPALSPAMGLTHGVHGIGDTVAISV
                HAAESAVSDIDAYMRLLDAALQ"
gene            complement(9978. .10583)
                /gene="Rv1126c"
CDS             complement(9978. .10583)
                /gene="Rv1126c"
                /note="Rv1126c, (MTCY22G8.15c), len: 201. Similar in
                N-terminus to M. leprae hypothetical  protein TR:O05567
                MLCB33.17 (141 aa), fasta scores; opt: 332 z-score: 552.7
                E(): 1.4e-23, 58.4% identity in 101 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv1126c"
                /protein_id="CAB09058.1"
                /db_xref="PID:e317246"
                /db_xref="PID:g2117220"
                /db_xref="GI:2117220"
                /db_xref="SPTREMBL:O06578"
```

```
                /translation="MSELTVLQAVRLKGRVITTDLAQTLGEDLADVAATVDRLTAAGL
                LVDATPLRISPSGRMRLDDLLAEERNRADSTVLAAAYRDFRSVNADFKRLVTDWQLKG
                EKPNTHDDAEYDAAVLSRLDGVHRRVGPIIGTVAMQLPRLSRYPVKLRAALDKVKAGD
                IAWLTRPLIDSYHTVWFELHEELIQAVGLTRDEAAKSGDAQ"
gene            complement(10580. .12052)
                /gene="ppdK"
CDS             complement(10580. .12052)
                /gene="ppdK"
                /note="Rv1127c, (MTCY22G8.16c), len: 490. Highly similar
                to N-terminus of eg. PODK_BACSY P22983 pyruvate, phosphate
                dikinase (839 aa), fasta scores; opt: 786 z-score: 880.2
                E():0, 37.0% identity in 514 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="ppdK"
                /protein_id="CAB09038.1"
                /db_xref="PID:e317147"
                /db_xref="PID:g2117221"
                /db_xref="GI:2117221"
                /db_xref="SPTREMBL:O06579"
                /translation="MTRITRANGCPDGTLENAVVALDGGANYPREILGNKGHGIDMMR
                RHHLPVPPAFCITTEVGVRYLAAPGSTIAAIWDDVLDRMSWLETETSCTFGRGPNPLL
                VSVRSGATQSMPGMMDTILDVGMTDAVERVLARPGAADFAHDTRRRFTSMYRRIVGSA
                GPITDDPYAQLRASIEAVFASWNSPRAVAYRDHHGLDDQGGTAVVVQAMVFGNLTANS
                GAGVLSSRNPITGANEPFGEWLPGGQGDDVVSGLVAVAPITALRDQQPAVYDQLMAAA
                RSLERMAGDVQEIEFTVEDSQLWLLQTRGAERSAQAAVRLALQLHHEGLIDDTETLRR
                VTPTHIETLLRPSLQTETRLAAPLLAKGLPACPGVVSGTAYTEVDEALDAADRGEPVI
                LVRDHTRPEDVMGMLAAQGIVTEVGGAASHAAVVSRELGRVAVVGCGPGVAAALAGKE
                ITVDGYEGEVRQGVLALSAWSESDTPELRELADIAQRISS"
RBS             complement(12060. .12063)
                /note="possible RBS upstream of Rv1127c"
gene            complement(12265. .13620)
                /gene="Rv1128c"
CDS             complement(12265. .13620)
                /gene="Rv1128c"
                /note="Rv1128c, (MTCY22G8.17c), len: 451. Unknown ORF in
                REP13E12 degenerate repeat, highly similar to several
                M.tuberculosis proteins in REP13E12 repeats eg. Q50655
                (317 aa),fasta scores; opt: 1329 z-score: 1558.9 E(): 0,
                61.1% identity in 316 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv1128c"
                /protein_id="CAB09039.1"
                /db_xref="PID:e317148"
                /db_xref="PID:g2117222"
                /db_xref="GI:2117222"
                /db_xref="SPTREMBL:O06580"
                /translation="MCSTREEITEAFASLATALSRVLGLTFDALTTPERLALLEHCET
                ARRQLPSVEHTLINQIGEQSTEEELGGKLGLTLADRLRITRSEAKRRVAEAADLGQRR
                ALTGEPLPPLLTATAKAQRHGLIGDGHVEVIRAFVHRLPSWVDLKTLEKAERDLAKQA
                TQYRPDQLAKLAARIMDCLNPDGDYTDEDRARRRGLTLGKQDVDGMSRLSGYVTPELR
                ATIEAVWAKLAAPGMCNPEQKAPCVNGAPSKEQARRDTRSCPQRNHDALNAELRSLLT
                SGNLGQHNGLPASIIVTTTLKDLEAAAGAGLTGGGTILPISDVIRLARHANHYLAIFD
                RGKALALYHTKRLASPAQRIMLYAKDSGCSAPGCDVPGYYCEVHHVTPYAQCRNTDVN
                DLTLGCGGHHPLAERGWTTRKNAHGDTEWLPPPHLDHGQPRVNTFHHPEKLLADDEGD
                P"
repeat_unit     12269. .13593
                /gene="Rv1128c"
                /note="REP13E12, degenerate repeat (66% identical to
                REP13E12)"
gene            complement(13722. .15182)
                /gene="Rv1129c"
CDS             complement(13722. .15182)
                /gene="Rv1129c"
                /note="Rv1129c, (MTCY22G8.18c), len: 486. Unknown,
                possible regulator, 47.4% identity in 468 aa overlap with
                M. tuberculosis Rv0465c(MTV038.09c)(474 aa) E(): 0. Helix
                turn helix motif present from aa 32-53"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv1129c"
                /protein_id="CAB09040.1"
                /db_xref="PID:e317149"
                /db_xref="PID:g2117223"
                /db_xref="GI:2117223"
                /db_xref="SPTREMBL:O06581"
                /translation="MTRSNVLPVARTYSRTFSGARLRRLRQERGLTQVALAKALDLST
                SYVNQLENDQRPITVPVLLLLTERFDLSAQYFSSDSDARLVADLSDVFTDIGVEHAVS
                GAQIEEFVARMPEVGHSLVAVHRRLRAATEELEGYRSRATAETELPPARPMPFEEVRD
                FFYDRNNYIHDLDMAAERMFTESGMRTGGLDIQLAELMRDRFGISVVIDDNLPDTAKR
                RYHPDTKVLRVAHWLMPGQRAFQIATQLALVGQSDLISSIVATDDQLSTEARGVARIG
                LANYFAGAFLLPYREFHRAAEQLRYDIDLLGRRFGVGFETVCHRLSTLQRPRQRGIPF
                IFVRTDKAGNISKRQSATAFHFSRVGGSCPLWVVHDAFAQPERIVRQVAQMPDGRSYF
                WVAKTTAADGLGYLGPHKNFAVGLGCDLAHAHKLVYSTGVVLDDPSTEVPIGAGCKIC
                NRTSCAQRAFPYLGGRVAVDENAGSSLPYSSTEQSV"
gene            15203. .16783
                /gene="Rv1130"
CDS             15203. .16783
                /gene="Rv1130"
                /note="Rv1130, (MTCY22G8.19), len: 526. Some similarity
                toeg.PRPD_ECOLI P77243 prpd protein. escherichia coli
                (483aa), fasta scores; opt:  234 z-score: 273.3 E():
                3.3e-08, 27.0% identity in 429 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv1130"
                /protein_id="CAB09041.1"
                /db_xref="PID:e317150"
                /db_xref="PID:g2117224"
                /db_xref="GI:2117224"
                /db_xref="SPTREMBL:O06582"
                /translation="MPDQDTKVRFFRVFCWCPVLRMVRIMLMHAVRAWRSADDFPCTE
                HMAYKIAQVAADPVDVDPEVADMVCNRIIDNAAVSAASMVRRPVTVARHQALAHPVRH
                GAKVFGVEGSYSADWAAWANGVAARELDFHDTFLAADYSHPADNIPPLVAVAQQLGVC
                GAELIRGLVTAYEIHIDLTRGICLHEHKIDHVAHLGPAVAAGIGTMLRLDQETIYHAI
                GQALHLTTSTRQSRKGAISSWKAFAPAHAGKVGIEAVDRAMRGEGSPAPIWEGEDGVI
                AWLLAGPEHTYRVPLPAPGEPKRAILDSYTKQHSAEYQSQAPIDLACRLRERIGDLDQ
                IASIVLHTSHHTHVVIGTGSGDPQKFDPDASRETLDHSLPYIFAVALQDGCWHHERSY
                APERARRSDTVALWHKISTVEDPEWTRRYHCADPAKKAFGARAEVTLHSGEVIVDELA
                VADAHPLGTRPFERKQYVEKFTELADGVVEPVEQQRFLAVVESLADLESGAVGGLNVL
                VDPRVLDKAPVIPPGIFR"
RBS             16770. .16772
                /gene="Rv1130"
                /note="possible RBS upstream of Rv0896"
gene            16780. .17961
                /gene="gltA1"
CDS             16780. .17961
                /gene="gltA1"
                /note="Rv1131, (MTCY22G8.20), gltA1, len: 393. Probable
                citrate synthase, highly similar to eg.CISY_MYCSM P26491
                citrate synthase (ec4.1.3.7) (375 aa), fasta scores; opt:
                1942z-score: 2334.7E(): 0, 80.0% identity in 375 aa
                overlap; contains PS00480Citrate synthase signature.
                Similar to two other M. tuberculosis citrate synthases
                MTCY31.24 (33.1% identity in 381 aa overlap) and
                MTCY31.17c (31.8% identity in 371 aa overlap)"
                /codon_start=1
                /transl_table=11
                /product="gltA1"
                /protein_id="CAB09042.1"
                /db_xref="PID:e317151"
                /db_xref="PID:g2117225"
                /db_xref="GI:2117225"
                /db_xref="SPTREMBL:O08395"
                /translation="MTGPLAAARSVAATKSMTAPTVDERPDIKKGLAGVVVDTTAISK
                VVPQTNSLTYRGYPVQDLAARCSFEQVAFLLWRGELPTDAELALFSQRERASRRVDRS
                MLSLLAKLPDNCHPMDVVRTAISYLGAEDPDEDDAAANRAKAMRMMAVLPTIVAIDMR
                RRRGLPPIAPHSGLGYAQNFLHMCFGEVPETAVVSAFEQSMILYAEHGFNASTFAARV
                VTSTQSDIYSAVTGAIGALKGRLHGGANEAVMHDMIEIGDPANAREWLRAKLARKEKI
                MGFGHRVYRHGDSRVPTMKRALERVGTVRDGQRWLDIYQVLAAEMASATGILPNLDFP
                TGPAYYLMGFDIASFTPIFVMSRITGWTAHIMEQATANALIRPLSAYCGHEQRVLPGT
                F"
misc_feature    17611. .17649
                /gene="gltA1"
                /note="PS00480 Citrate synthase signature"
```

```
     gene            17973. .19703
                     /gene="Rv1132"
     CDS             17973. .19703
                     /gene="Rv1132"
                     /note="Rv1132, (MTCY22G8.21), len: 576. Possible
                     membraneprotein, similar to M. tuberculosis protein
                     TR:O06827 MTCY493.23C (589 aa), fasta scores; opt: 1811,
                     E(): 0, 48.2% identity in 585 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein Rv1132"
                     /protein_id="CAB09043.1"
                     /db_xref="PID:e317152"
                     /db_xref="PID:g2117226"
                     /db_xref="GI:2117226"
                     /db_xref="SPTREMBL:O06583"
                     /translation="MGFLQPRLPDIDLAEWSQGSRSQKIRPMAQHWAEVGFGTPVLLH
                     LFYVAKILLYVLVGWLIVLTTKGIDGFTDAAAWYAEPIVFEKVVLYTMLFEVIGLGCG
                     FGPLNNRFFPPMGSILYWMRFGTIRLPPWPDRVPWTRGTKRKPVDVALYALLVMMLLS
                     ALFTDGAGPIPELGTTVGLLPAWQIVLILLLLGVLGLRDKVIFLAARGEVYATLTVTF
                     LFGRLNGIDMIVAAKLVFLVIWIGAATSKLNRHFPFVISTMMSNNPLFRPRFIKRMFF
                     KKFPGDLRPGLLSRIVAHVSTVIEMCVPVVLFVAHGGWPTVVAATIMVCFHLGILTAI
                     PMGVPLEWNVFMIFGVLSLFVGHACLGLADVKNPVPLAILIAVVAGIVIAGNVFPRKI
                     SFLAAMRYYAGNWDTTLWCIKPSAEDKINRGIVAIASMPAAQLERFYGKDRAQIPMYL
                     GYAFRAMNSHGRALFTLAHRAMAGHDEDDYVITDGERVCSTAVGWNFGDGHLHNEQLI
                     AAMQQRCGFQPGEVRVVLLDAQPIHRQTQEYRLVDAATGEFERGYVRVADMVNRQPWD
                     DDVPVHVLPG"
     gene            complement(19715. .21994)
                     /gene="metE"
     CDS             complement(19715. .21994)
                     /gene="metE"
                     /note="Rv1133c, (MTC22G8.22), metE, len: 759. Probable
                     5-methyltetrahydropteroyltriglutamate--homocysteine
                     methyltransferase (methionine synthase, vitamin-B12
                     independent isozyme), highly similar to eg. METE_ECOLI
                     P25665 5-methyltetrahydropteroyltriglutamate--homocysteine
                     methyltransferase (752 aa), fasta scores; opt: 2251
                     z-score: 2538.0 E(): 0, 48.1% identity in 756aa overlap;
                     start site chosen by homology"
                     /codon_start=1
                     /transl_table=11
                     /product="metE"
                     /protein_id="CAB09044.1"
                     /db_xref="PID:e317153"
                     /db_xref="PID:g2117227"
                     /db_xref="GI:2117227"
                     /db_xref="SWISS-PROT:O06584"
                     /translation="MTQPVRRQPFTATITGSPRIGPRRELKRATEGYWAGRTSRSELE
                     AVAATLRRDTWSALAAAGLDSVPVNTFSYYDQMLDTAVLLGALPPRVSPVSDGLDRYF
                     AAARGTDQIAPLEMTKWFDTNYHYLVPEIGPSTTFTLHPGKVLAELKEALGQGIPARP
                     VIIGPITFLLLSKAVDGAGAPIERLEELVPVYSELLSLLADGGAQWVQFDEPALVTDL
                     SPDAPALAEAVYTALCSVSNRPAIYVATYFGDPGAALPALARTPVEAIGVDLVAGADT
                     SVAGVPELAGKTLVAGVVDGRNVWRTDLEAALGTLATLLGSAATVAVSTSCSTLHVPY
                     SLEPETDLDDALRSWLAFGAEKVREVVVLARALRDGHDAVADEIASSRAAIASRKRDP
                     RLHNGQIRARIEAIVASGAHRGNAAQRRASQDARLHLPPLPTTTIGSYPQTSAIRVAR
                     AALRAGEIDEAEYVRRMRQEITEVIALQERLGLDVLVHGEPERNDMVQYFAEQLAGFF
                     ATQNGWVQSYGSRCVRPPILYGDVSRPRAMTVEWITYAQSLTDKPVKGMLTGPVTILA
                     WSFVRDDQPLADTANQVALAIRDETVDLQSAGIAVIQVDEPALRELLPLRRADQAEYL
                     RWAVGAFRLATSGVSDATQIHTHLCYSEFGEVIGAIADLDADVTSIEAARSHMEVLDD
                     LNAIGFANGVGPGVYDIHSPRVPSAEEMADSLRAALRAVPAERLWVNPDCGLKTRNVD
                     EVTASLHNMVAAAREVRAG"
     RBS             complement(22001. .22005)
                     /note="posssible RBS upstream of Rv1133c"
     source          22450. .>22550
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37Rv"
                     /db_xref="taxon:1773"
                     /clone="I65"
BASE COUNT     3693 a   7255 c   7548 g   4054 t
ORIGIN
```

Query Match 4.0%; Score 39.4; DB 1; Length 22550;

```
  Best Local Similarity   45.6%;  Pred. No. 4.1;
  Matches  139;  Conservative    0;  Mismatches  166;  Indels    0;  Gaps    0;

Qy      22 GCGGGCATCAGCTCCCTTGACCCAGTGGATATCGGTGGCCCCGTTATTCGTCCAGGTGCC 81
           || ||||    |  | | ||  | || |||   |     | ||      |  ||   ||
Db   15941 GCCGGCAAGGTCGGCATCGAGGCGGTCGATCGGGCGATGCGCGGCGAGGGCTCACCGGCT 16000

Qy      82 CAGGGAGGAGGACCCGCCTGCAGCATGAACCTGTGGCTCCTGGCCTGCCTGGTGGCCGGC 141
           | |    | |    ||    | |  ||| |   ||||| |||||| | |  | |  |  |
Db   16001 CCGATCTGGGAGGGCGAGGACGGGGTGATCGCCTGGCTGCTGGCCGGACCCGAGCACACC 16060

Qy     142 TTCCTGGGAGCCTGGGCCCCCGCTGTCCACGCCCAAGGTGTCTTTGAGGACTGCTGCCTG 201
           | || ||   | | | || |  |||   |  || |  | | | ||  |||| ||| |
Db   16061 TACCGGGTGCCGTTGCCCGCACCTGGTGAACCCAAGCGCGCCATTCTGGACAGCTACACC 16120

Qy     202 GCCTACCACTACCCCATTGGGTGGGCTGTGCTCCGGCGCGCCTGGACTTACCGGATCCAG 261
             | |||| | |      |     |   | |  || |  | | | |||| | |
Db   16121 AAGCAACACTCCGCGGAGTACCAGAGCCAGGCGCCGATCGACCTGGCCTGCCGGCTACGT 16180

Qy     262 GAGGTGAGCGGGAGCTGCAATCTGCCTGCTGCGATATTCTACCTCCCCAAGAGACACAGG 321
           |||   | |||      | | | |   ||  |||| |   |  | | |||     ||||
Db   16181 GAGCGTATCGGCGATCTCGACCAGATCGCGTCGATCGTGCTGCACACCAGCCACCACACC 16240

Qy     322 AAGGT 326
            | ||
Db   16241 CATGT 16245


RESULT   9
HSU85768
LOCUS       HSU85768      360 bp    mRNA            PRI       01-APR-1997
DEFINITION  Human myeloid progenitor inhibitory factor-1 MPIF-2 mRNA, complete
            cds.
ACCESSION   U85768
NID         g1916251
VERSION     U85768.1  GI:1916251
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Patel,V.P., Kreider,B.L., Li,Y., Li,H., Leung,K., Salcedo,T.,
            Nardelli,B., Pippalla,V., Gentz,S., Thotakura,R., Parmelee,D.,
            Gentz,R. and Garotta,G.
  TITLE     Molecular and functional characterization of two novel human C-C
            chemokines as inhibitors of two distinct classes of myeloid
            progenitors
  JOURNAL   J. Exp. Med. (1997) In press
REFERENCE   2  (bases 1 to 360)
  AUTHORS   Li,H. and Patel,V.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JAN-1997) Cell Biology, Human Genome Sciences, 9410
            Keywest Ave., Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .360
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             1. .360
                     /note="myeloid progenitor inhibitory factor-2"
                     /codon_start=1
                     /product="MPIF-2"
                     /protein_id="AAB51135.1"
                     /db_xref="PID:g1916252"
                     /db_xref="GI:1916252"
                     /translation="MAGLMTIVTSLLFLGVCAHHIIPTGSVVIPSPCCMFFVSKRIPE
                     NRVVSYQLSSRSTCLKGGVIFTTKKGQQFCGDPKQEWVQRYMKNLDAKQKKASPRARA
                     VAVKGPVQRYPGNQTTC"
BASE COUNT       85 a    106 c     96 g     73 t
ORIGIN
```

Query Match 3.9%; Score 38.6; DB 10; Length 360;

```
  Best Local Similarity   55.6%;  Pred. No. 6.8;
  Matches   74;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy     258 CCAGGAGGTGAGCGGGAGCTGCAATCTGCCTGCTGCGATATTCTACCTCCCCAAGAGACA 317
           ||||  |   ||| |||||      ||    |  |   |  ||| |  | ||||||
Db     150 CCAGCTGTCCAGCAGGAGCACATGCCTCAAGGGAGGAGTGATCTTCACCACCAAGAAGGG 209

Qy     318 CAGGAAGGTGTGTGGGAACCCCAAAAGCAGGGAGGTGCAGAGAGCCATGAAGCTCCTGGA 377
           |  | || | |||||  ||||||       |  ||| |||||   |||||||  ||||||
Db     210 CCAGCAGTTCTGTGGCGACCCCAAGCAGGAGTGGGTCCAGAGGTACATGAAGAACCTGGA 269

Qy     378 TGCTCGAAATAAG 390
           ||     | |||
Db     270 CGCCAAGCAGAAG 282


RESULT  10
MTV017
LOCUS       MTV017      67200 bp    DNA             BCT       23-JUN-1998
DEFINITION  Mycobacterium tuberculosis H37Rv complete genome; segment 48/162.
ACCESSION   AL021897 AL123456
NID         g3256022
VERSION     AL021897.1  GI:3256022
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium.
REFERENCE   1  (bases 1 to 67200)
  AUTHORS   Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
            Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
            Tekaia,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
            Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S.,
            Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
            Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
            Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
            Squares,S., Sqares,R., Sulston,J.E., Taylor,K., Whitehead,S. and
            Barrell,B.G.
  TITLE     Deciphering the biology of Mycobacterium tuberculosis from the
            complete genome sequence
  JOURNAL   Nature 393 (6685), 537-544 (1998)
  MEDLINE   98295987
  REMARK    Erratum:[[published erratum appears in Nature 1998 Nov
            12;396(6707):190]]
REFERENCE   2  (bases 1 to 67200)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
            tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
            Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
            Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
            75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
COMMENT     On Jun 26, 1998 this sequence version replaced gi:2896684.
            Notes:
            Details of M. tuberculosis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
            been renumbered from the original cosmid submissions but the old
            gene designations are in brackets after the new gene numbers.
            Gene prediction was based on a Hidden Markov Model of TB genes
            implemented in TBparse (Krogh) supplemented with visual inspection
            of positional base preference in codons, especially where there is
            an increase in the observed/expected third position G + C.
            CAUTION:  In some cases we may not have predicted the correct
            initiation codon.  Where possible we choose an initiation codon
            (atg, gtg, or ttg) which is preceded by an upstream ribosome
            binding site sequence (optimally 5-13bp before the initiation
            codon).  If this cannot be identified we choose the most upstream
            initiation codon.
FEATURES              Location/Qualifiers
     source           1. .67200
                      /organism="Mycobacterium tuberculosis"
                      /strain="H37Rv"
                      /db_xref="taxon:1773"
     misc_feature     <1. .66170
                      /note="fragment designated v017. Does not represent a
                      physical clone"
     RBS              178. .181
                      /note="possible RBS, ggag, for Rv1049"
     gene             184. .630
                      /gene="Rv1049"
     CDS              184. .630
                      /gene="Rv1049"
                      /note="Rv1049, (MTV017.02), len: 148. Probable repressor
                      similar to many e.g. P74870 (149 aa). FASTA scores:
                      sptr|P74870|P74870NEGATIVE REGULATOR OF EMR LOCUS EMR (149
                      aa) opt:146 z-score: 200.3 E(): 0.0011;  31.6% identity in
                      95 aaoverlap. TBparse score is 0.892. Contains probable
                      helix-turn-helix motif at aa 58-79 (Score 1495, +4.28 SD)"
                      /codon_start=1
                      /transl_table=11
                      /product="hypothetical protein Rv1049"
                      /protein_id="CAA17165.1"
                      /db_xref="PID:e1251928"
                      /db_xref="PID:g2896686"
                      /db_xref="GI:2896686"
                      /db_xref="SPTREMBL:O53397"
                      /translation="MGKGAAFDECACYTTRRAARQLGQAYDRALRPSGLTNTQFSTLA
                      VISLSEGSAGIDLTMSELAARIGVERTTLTRNLEVMRRDGLVRVMAGADARCKRIELT
                      AKGRAALQKAVPLWRGVQAEVTASVGDWPRVRRDIANLGQAAEACR"
     RBS              668. .672
                      /note="possible RBS, aggag, for Rv1050"
     gene             679. .1584
                      /gene="Rv1050"
     CDS              679. .1584
                      /gene="Rv1050"
                      /note="Rv1050, (MTV017.03), len: 301. Probable
                      oxidoreductase similar to many e.g. MTCY48.22C(341 aa).
                      FASTA scores:sp|Q10783|Y04M_MYCTU PUTATIVE OXIDOREDUCTASE
                      CY48.22C  (341aa) opt:462 z-score: 533.9 E():  3e-22;
                      33.6% identity in265 aa overlap. TBparse score is 0.896"
                      /codon_start=1
                      /transl_table=11
                      /product="hypothetical protein Rv1050"
                      /protein_id="CAA17166.1"
                      /db_xref="PID:e1251929"
                      /db_xref="PID:g2896687"
                      /db_xref="GI:2896687"
                      /db_xref="SPTREMBL:O53398"
                      /translation="MARQRFRDQVVLITGASSGIGEATAKAFAREGAVVALAARREGA
                      LRRVAREIEAAGGRAMVAPLDVSSSESVRAMVADVVGEFGRIDVVFNNAGVSLVGPVD
                      AETFLDDTREMLEIDYLGTVRVVREVLPIMKQQRSGRIMNMSSVVGRKAFARFAGYSS
                      AMHAIAGFSDALRQELRGSGIAVSVIHPALTQTPLLANVDPADMPPPFRSLTPIPVHW
                      VAAAVLDGVARRRARVVVPFQPRLLMVGDAFSPRYGDRVVRLLESKIFGRLIGSYRGS
                      VYRHQPTESAKAQAAQPERGYSSAR"
     gene             complement(1743. .2498)
                      /gene="Rv1051c"
     CDS              complement(1743. .2498)
                      /gene="Rv1051c"
                      /note="Rv1051c, (MTV017.04c), len: 251. Unknown but
                      similar to protein encoded by Lactococcus lactis plasmid
                      pNP40, LLU36837_2 (298 aa). FASTA scores:
                      U36837|LLU36837_2 Lactococcus lactisplasmid pNP40, (298
                      aa) opt:  194 z-score: 245.5E(): 3.5e-06; 30.3% identity
                      in 155 aa overlap. TBparse score is 0.912. Contains
                      possible helix-turn-helix motif at aa197-218 (Score 1097,
                      +2.92 SD)"
                      /codon_start=1
                      /transl_table=11
                      /product="hypothetical protein Rv1051c"
                      /protein_id="CAA17167.1"
                      /db_xref="PID:e1251930"
                      /db_xref="PID:g2896688"
                      /db_xref="GI:2896688"
                      /db_xref="SPTREMBL:O53399"
```

/translation="MRADVTAEHLTQVVRDIAVIDIDDGVAFNLDTSSVQEIRERADY
PGLRVRVAMSVGPWQGIAAWDVSTGEPIAPWPTRVTIDRILGEPITLLGYAPETIIAE
KGVTILERGITSTRWRDYVDIVQLDRRGIDDDELLRSARAVAQYRGATLEPVAPHLAG
YGAVAQAKWATEHGRCQHCWRHWKPAHVGRRNMDLLDAKQVSEMIGVPVGTLRHWRHS
DIGPASFTLGRRVVYRRDEVSRWISKRESATRR"

RBS 3507. .3513
/note="possible RBS, aaggtgg, for Rv1052"

gene 3521. .3910
/gene="Rv1052"

CDS 3521. .3910
/gene="Rv1052"
/note="Rv1052, (MTV017.05), len: 129. Unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1052"
/protein_id="CAA17168.1"
/db_xref="PID:e1251931"
/db_xref="PID:g2896689"
/db_xref="GI:2896689"
/db_xref="SPTREMBL:O53400"
/translation="MDCCEERGVARHKGLSQVGTPGCPRWSQAVSCRCSAYREAAVTA
VQMPLTPGYGETPLPHDELAALLPEVVEVLDKPITRADVYDLEQGLQDQVFDLLMPTA
VEGSLSLDELLSDHFVRDLHARMFGPV"

gene complement(3809. .4084)
/gene="Rv1053c"

CDS complement(3809. .4084)
/gene="Rv1053c"
/note="Rv1053c, (MTV017.06c), len: 91. Unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1053c"
/protein_id="CAA17169.1"
/db_xref="PID:e1251932"
/db_xref="PID:g2896690"
/db_xref="GI:2896690"
/db_xref="SPTREMBL:O53401"
/translation="MDSHKVCMNNNTQLPTGPIIGVHPAVRDGVERVAYLDGDLLRCN
TDVEFTSSPPPGPVLYRTKHTRVEIADEMVTEKLIKRQRAFNSRRHQ"

gene 4726. .4932
/gene="Rv1054"

CDS 4726. .4932
/gene="Rv1054"
/note="Rv1054, (MTV017.07), len: 68. Unknown but similar to hypothetical Mycobacterium tuberculosis protein MTCY3G12_25 (151 aa shows similarity to integrases) and to Mycobacterium paratuberculosis integrase MSGINT_1 (191 aa). This orfcontinuesin another frame as MTV017.08 but no error can befound toaccount for frameshift. FASTA scores: Z79702|MTCY3G12_25 (151 aa) opt: 273 z-score: 364.1 E(): 8.8e-13; 64.7% identity in 68 aa overlap; and L39071|MSGINT_1 (191aa)opt: 105 z-score: 148.5 E(): 0.9; 31.8% identity in85 aaoverlap."
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1054"
/protein_id="CAA17170.1"
/db_xref="PID:e1251933"
/db_xref="PID:g2896691"
/db_xref="GI:2896691"
/db_xref="SPTREMBL:O53402"
/translation="MTGKGIVESTTKTKRDRHVPVPEPVWRRLHAELPTDPNALVFPG
RKGGFLPLGEYRWAFDNAGDQVGIE"

gene 4935. .5171
/gene="Rv1055"

CDS 4935. .5171
/gene="Rv1055"
/note="Rv1055, (MTV017.08), len: 78. Partial orf, first 49aa similar to hypothetical Mycobacterium tuberculosis protein MTCY3G12_25 (151 aa shows similarity to integrases) andto Mycobacterium paratuberculosis integrase L39071|MSGINT_1 (191 aa)and to many other integrases or transposases. FASTAscores:Z79702|MTCY3G12_25 (151 aa) opt: 291 z-score:428.6 E():2.2e-16; 74.3% identity in 70 aa overlap; andgp|L39071|MSGINT_1 (191 aa) opt: 146 z-score: 221.0 E(): 8.3e-05; 52.1% identity in 48 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1055"
/protein_id="CAA17171.1"
/db_xref="PID:e1251934"
/db_xref="PID:g2896692"
/db_xref="GI:2896692"
/db_xref="SPTREMBL:O53403"
/translation="MVPHGLGHTTASLAISAGANVKVVQRLLGHAAAAMTLDRHGHLL
NDDLAVWPMRCAKSSRTLRYHCGMRRRNRVGLRA"

tRNA 5194. .5267
/note="leuX, tRNA-Leu : anticodon taa"

RBS 5413. .5418
/note="possible RBS, aaggag, for Rv1056"

gene 5426. .6190
/gene="Rv1056"

CDS 5426. .6190
/gene="Rv1056"
/note="Rv1056, (MTV017.09), len: 254. Unknown but some similarity in C-terminal part to Mycobacterium tuberculosis MTCI5.14 (126aa). FASTA scores: Z92770|MTCI5_14 Mycobacteriumtuberculosis cosmid I (126 aa) opt: 254 z-score: 325.8 E(): 1.2e-10;43.4% identity in 106 aa overlap. TBparse scoreis0.925"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1056"
/protein_id="CAA17172.1"
/db_xref="PID:e1251935"
/db_xref="PID:g2896693"
/db_xref="GI:2896693"
/db_xref="SPTREMBL:O53404"
/translation="MSVDYPQMAATRGRIEPAPRRVRGYLGHVLVFDTSAARYVWEVP
YYPQYYIPLADVRMEFLRDENHPQRVQLGPSRLHSLVSAGQTHRSAARVFDVDGDSPV
AGTVRFNWDPLRWFEEDEPIYGHPRNPYQRADALRSHRHVRVELDGIVLADTRSPVLL
FETGIPTRYYIDPADIAFEHLEPTSTQTLCPYKGTTSGYWSVRVGDAVHRDLAWTYHY
PLPAVAPIAGLVAFYNEKVDLTVDGVALPRPHTQFS"

RBS 7180. .7186
/note="possible RBS, aggagga, for Rv1057"

gene 7194. .8375
/gene="Rv1057"

CDS 7194. .8375
/gene="Rv1057"
/note="Rv1057, (MTV017.10), len: 393. Unknown but some similarity to surface antigen of Methanosarcina mazeii, MMSAG_1 (491 aa). FASTA scores: X84710|MMSAG_1 M.mazei surface antigen genes orf492 (491 aa) opt: 363 z-score: 402.6 E():6.2e-15; 31.3% identity in 294 aa overlap. TBparse score is0.892"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1057"
/protein_id="CAA17173.1"
/db_xref="PID:e1251936"
/db_xref="PID:g2896694"
/db_xref="GI:2896694"
/db_xref="SPTREMBL:O53405"
/translation="MSVMNGREVARESRDAQVFEFGTAPGSAVVKIPVQGGPIGGIAI
SRDGSLLVVTNNGTDTVSVVGTDTCRVTQTVTSVNEPFAIAMGNAEANRAYVSTVSSA
YDAIAVIDVATNTVLGTHPLALSVSDLTLSPDDKYLYVSRNGTRGADVAVLDTTTGAL
IDVVDVSQAPGTTTQCVRMSPDGSVLYVGANGPSGGLLVVITTRAQSDGGRIGSRSRS
RQKSSKPRGNQAAAGLRVVATIDIGSSVRDVALSPDGAIAYVASCGSDFGAVVDVIDT
RTHQITSSRAISEIGGLVTRVSVSGDADRAYLVSEDRVTVLCTRTHDVIGTIRTGQPS
CVVESPDGKYLYIADYSGTITRTAVASTIVSGTEQLALQRRGSMQWFSPELQQYAPAL
A"

RBS 8468. .8474
/note="possible RBS, aaggcgg, for Rv1058"

gene 8482. .10113
/gene="fadD14"

CDS 8482. .10113

/gene="fadD14"
/note="Rv1058, (MTV017.11), len: 543. Probable fatty acidCoA-ligase similar to MEDIUM-CHAIN-FATTY-ACID--COA LIGASE (546 aa) (EC6.2.1.-) and 4-COUMARATE-COA LIGASE (EC 6.2.1.12). ContainsPS00455 Putative AMP-binding domain signature. FASTA scores: sp|Q00594|ALKK_PSEOL MEDIUM-CHAIN-FATTY-ACID--COA LIGA(546 aa) opt: 1468 z-score: 1731.7 E(): 0; 41.1%identity in538 aa overlap. TBparse score is 0.916"
/codon_start=1
/transl_table=11
/product="fadD14"
/protein_id="CAA17174.1"
/db_xref="PID:e1251937"
/db_xref="PID:g2896695"
/db_xref="GI:2896695"
/db_xref="SPTREMBL:O53406"
/translation="MYGTMQDFPLTITAIMRHGCGVHGRRTVTTATGEGYRHSSYRDV GQRAGQLANALRRLGVTGDQRVATFMWNNTEHLVTYFAVPSMGAVLHTLNIRLFPEQI AYVTNEAEDRVILVDLSLARLLAPVLPKLDTVHTVIAVGEGDTTPLREAGKTVLRFAE LIDAESPDFGWPQIDENSAAAMCYTSGTTGNPKGVVYSHRSSFLHTMAACTTNGIGVG SSDKVLPIVPMFHANGWGLPYAALMAGADLVLPDRHLDARSLIHMVETLKPTLAGAVP TIWNDVMHYLEKDPDHDMSSLRLVACGGSAVPESLMRTFEDKHDVQIRQLWGMTETSP LATMAWPPPGTPDDQHWAFRITQGQPVCGVETRIVDDDGQVLPNDGNAVGEVEVRGPW IAGSYYGGRDESKFDSGWLRTGDVGRIDEQGFITLTDRAKDVIKSGGEWISSVELENC LIAHPDVLEAAVVGVPDERWQERPLAVVVVREGATVSAGDLRAFLADKVVRWWLPERW AFVDEIPRTSVGKYDKKAIRSRYAEGAYQITEVHT"
misc_feature 9025. .9060
/gene="fadD14"
/note="PS00455 Putative AMP-binding domain signature"
misc_feature 10115. .10188
/note="Conserved intergenic region. Similar to positions (+1) 25900 25973 of MTY20B11 and (+1) 19989 20052 of MTCY01B2"
gene 10189. .11253
/gene="Rv1059"
CDS 10189. .11253
/gene="Rv1059"
/note="Rv1059, (MTV017.12), len: 354. Unknown but similarto Mycobacterium tuberculosis MTCY21C12.20c (358 aa). FASTAscores: Z95210|MTY21C12_20 Mycobacterium tuberculosis cosmi(358 aa)opt: 338 z-score: 396.5 E(): 1.4e-14; 33.1% identity in 363 aa overlap. TBparse score is 0.909"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1059"
/protein_id="CAA17175.1"
/db_xref="PID:e1251938"
/db_xref="PID:g2896696"
/db_xref="GI:2896696"
/db_xref="SPTREMBL:O53407"
/translation="MTMSLRVIQWATGSVGVAAIKGVLQHPELELVGCWVHSAAKSGK DVGEIIGSPPLGVIATNSIDDVLALDADAVIYAPLLPSVDEVAALLRSGKNVVTPLGW FYPSEKEAAPLEVAAQAGNATLHGAGIGPGAVTELFPLLLSVMSTGVTFVRSEEFSDL RSYGAPDVLRYVMGFGGTPDSALTGPMQKILDGGFLQSVRLCVDRLGFAADPQIRTSQ EVAVATAPIDSPIGVIEPGQVAGRRFHWEALVEDTVVVQIAVNWLMGSENLDPPWSFG PAGERYEIEVRGSPDTCVTIKGWQPQTVAAGLKSNPGIVATAAHCVNAIPATCAAPAG IQSFFDLPLITGRAAPGLAR"
gene 11306. .11779
/gene="Rv1060"
CDS 11306. .11779
/gene="Rv1060"
/note="Rv1060, (MTV017.13), len: 157. Unknown. TBparse score is 0.912"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1060"
/protein_id="CAA17176.1"
/db_xref="PID:e1251939"
/db_xref="PID:g2896697"
/db_xref="GI:2896697"
/db_xref="SPTREMBL:O53408"
/translation="MAKSVVVEQSRAIPVQSEDAFGGTLAAALPVICSHWYGLIPPIK EVRDQTGAWDSVGQARVITMVGGGRVREELTSVDPPRSFGYTLTDIKGPLAPLVALVE GKWSFAPADTGTTVTWQWTIHPRSALAAPVLPVFARMWRGYARGVLEKLSALLVG"
gene 11813. .12676
/gene="Rv1061"
CDS 11813. .12676
/gene="Rv1061"
/note="Rv1061, (MTV017.14), len: 287. Unknown but similarto hypothetical proteins from various bacteria e.g. Synechocystis sp. PCC6803 D64002|SYCSLRD_75 (304 aa). FASTA scores: D64002|SYCSLRD_75 (304 aa) opt: 245 z-score: 307.8 E():1.2e-09; 27.1% identity in 258 aa overlap. TBparse score is0.923"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1061"
/protein_id="CAA17177.1"
/db_xref="PID:e1251940"
/db_xref="PID:g2896698"
/db_xref="GI:2896698"
/db_xref="SPTREMBL:O53409"
/translation="MCRLFGLHSGTDAVTATFWLLNASDSLAEQSRRNPDGTGLGVFD EHHQPRLHKQPIAAWQDADFATEAHELTGTTFVAHVRYATTGSLDIRNTHPFLQDGRI FAHNGVVEGLDVLDERLREVGADDLVLGQTDSERVFALITASIRARDGNESAGLIDAL RWLAANVPIYAVNVLLSTATDVWALRYPESHELYILDRRGDGAPEFHLRSKRIRAHST HLRERSSVVFATEPMDDNPRWRLLDAGELVHVDAALRVNRSLVLPDPPRHPIRREDLS EPVLHAQHTSA"
gene 12681. .13538
/gene="Rv1062"
CDS 12681. .13538
/gene="Rv1062"
/note="Rv1062, (MTV017.15), len: 285. Unknown but some similarity tohypothetical protein in Bacillus subtilis,BSUB0008_176 (260aa) . FASTA scores: Z99111|BSUB0008_176 (260 aa)opt: 163z-score: 199.1 E(): 0.0013; 27.4% identity in 179aaoverlap. TBparse score is 0.902"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1062"
/protein_id="CAA17178.1"
/db_xref="PID:e1251941"
/db_xref="PID:g2896699"
/db_xref="GI:2896699"
/db_xref="SPTREMBL:O53410"
/translation="MTTRRALVLAGGGLAGIAWETGVLRGIADESPAAARLLLDSDVL VGTSAGATVAAQISSGCPLDTLYERQLAETSAEIDPGVDIDAITDLFLTAVTEPHIST RRRLQRIGAVALAVDTVPESVRRQVIAQRLPSHDWPDRVLRVTAIDIATGELVVFHRE SNVALVDAVAASCSVPGAWPPVTIAGRRYMDGGVASSVNLGVADDCDAAVVLVPAGAD APSPFGGGAAAEIAAATGMVFAVFADDDSLAAFGPNPLDPLCRVNSAMAGRQQGRREA QAVARLLGV"
gene complement(13539. .14621)
/gene="Rv1063c"
CDS complement(13539. .14621)
/gene="Rv1063c"
/note="Rv1063c, (MTV017.16c), len: 360. Unknown but similar to sp|P37053|YCHK_ECOLI. FASTA scores: sp|P37053|YCHK_ECOLI HYPOTHETICAL 34.4 KD PROTEIN IN H (314 aa) opt: 487 z-score: 545.0 E(): 7.2e-23; 32.7% identity in 321 aa overlap. TBparsescore is 0.893. Also partially similar to Rv3239c(MTCY20B11.14c)"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1063c"
/protein_id="CAA17179.1"
/db_xref="PID:e1251942"
/db_xref="PID:g2896700"
/db_xref="GI:2896700"
/db_xref="SPTREMBL:O53411"
/translation="MPAPAALRVRGSSSPRVALALGSGGARGYAHIGVIQALRERGYD IVGIAGSSMGAVVGGVHAAGRLDEFAHWAKSLTQRTILRLLDPSISAAGILRAEKILD AVRDIVGPVAIEQLPIPYTAVATDLLAGKSVWFQRGPLDAAIRASIAIPGVIAPHEVD GRLLADGGILDPLPMAPIAGVNADLTIAVSLNGSEAGPARDAEPNVTAEWLNRMVRST

SALFDVSAARSLLDRPTARAVLSRFGAAAAESDSWSQAPEIEQRPAGPPADREEAADT
PGLPKMGSFEVMNRTIDIAQSALARHTLAGYPADLLIEVPRSTCRSLEFHRAVEVIAV
GRALATQALEAFEIDDDESAAATIEG"

gene complement(14702. .15121)
/gene="lpqV"

CDS complement(14702. .15121)
/gene="lpqV"
/note="Rv1064c, (MTV017.17c), len: 139. Unknown but probably lipoprotein. Has N-terminal signal sequence and appropriately positioned PS00013 Prokaryotic membrane lipoprotein lipid attachment site. TBparse score is 0.912"
/codon_start=1
/transl_table=11
/product="lpqV"
/protein_id="CAA17180.1"
/db_xref="PID:e1251943"
/db_xref="PID:g2896701"
/db_xref="GI:2896701"
/db_xref="SPTREMBL:O53412"
/translation="MRPSRYAPLLCAMVLALAWLSAVAGCSRGGSSKAGRSSSVAGTL
PAGVVGVSPAGVTTRVDAPAESTEEEYYQACHAARLWMDAQPGSGESLIEPYLAVVQA
SPSGVAGSWHIRWAALTPARQAAVIVAARAAANAECG"

misc_feature complement(15044. .15076)
/gene="lpqV"
/note="PS00013 Prokaryotic membrane lipoprotein lipid attachment site"

RBS 15222. .15225
/note="possible RBS, ggag, for Rv1065"

gene 15233. .15799
/gene="Rv1065"

CDS 15233. .15799
/gene="Rv1065"
/note="Rv1065, (MTV017.18), len: 188. Unknown but some similarity to hypothetical Streptomyces coelicolor protein AL0209|SC4H8_11 (182 aa). FASTA scores: gp|AL0209|SC4H8_11 (182 aa) opt: 156 z-score: 200.7 E(): 0.0011; 31.3% identity in 195 aaoverlap. TBparse score is 0.921"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1065"
/protein_id="CAA17181.1"
/db_xref="PID:e1251944"
/db_xref="PID:g2896702"
/db_xref="GI:2896702"
/db_xref="SPTREMBL:O53413"
/translation="MVMPLVTPTTAVPSPGPTRLRVADLLRATDQAADDVLGGRCDHL
LPDGGVPQTQRWYTRIHGDEELDIWLISWVPGQPTELHDHGGSLGALTVLSGSLNEYR
WDGRRLRRRRLDAGDQAGFPLGWVHDVVWAPRPIGGPDAAGMAVAPTLSVHAYSPPLT
AMSYYEITERNTLRRQRTELTDQPEGSG"

RBS 15785. .15789
/gene="Rv1065"
/note="possible RBS, gaagg, for Rv1066"

gene 15796. .16191
/gene="Rv1066"

CDS 15796. .16191
/gene="Rv1066"
/note="Rv1066, (MTV017.19), len: 131. Unknown but strong similarity to hypothetical Streptomyces coelicolor protein AL0209|SC4H8_10 (132 aa). FASTA scores: gp|AL0209|SC4H8_10 (132 aa) opt: 429 z-score: 547.6 E(): 5.2e-23; 57.1% identity in 119 aa overlap. TBparse score is 0.859"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1066"
/protein_id="CAA17182.1"
/db_xref="PID:e1251945"
/db_xref="PID:g2896703"
/db_xref="GI:2896703"
/db_xref="SPTREMBL:O53414"
/translation="MSRIDRVLEAARRRYRRLAADQVPEAARRGAVLVDIRPQAQRAR
EGEVPGALVIERNVLEWRCDPTSDARLPQAVDDDVEWVILCSEGYTSSLAAASLLDLG
LHRATDVVGGYRALAAGGVLAELGGAVGG"

gene complement(16219. .46828)
/gene="PE_PGRS"

CDS complement(16219. .18222)
/gene="PE_PGRS"
/note="Rv1067c, (MTV017.20c), len: 667. Member of M. tuberculosis PE-family, Gly-, Ala-rich PGRS subfamily. ContainsPS00583 pfkB family of carbohydrate kinases signature 1, probably fortuitous. FASTA scores: AL0091|MTV004_44 (731 aa)opt: 2227 z-score: 1261.4 E(): 0; 55.6% identity in 710aaoverlap. TBparse score is 0.837"
/codon_start=1
/transl_table=11
/product="PE_PGRS"
/protein_id="CAA17183.1"
/db_xref="PID:e1251946"
/db_xref="PID:g2896704"
/db_xref="GI:2896704"
/db_xref="SPTREMBL:O53415"
/translation="MSFVLVSPSQLMAAAADVAGIGSAISAANAAALAPTSVLAAAGA
DEVSAAVAALFSAHAGQYQQLGARAALFHEQFVQALTGAASAYASAEATNVEQQVLGL
INAPTQALLGRPLIGNGADGTAANPNGGAGGLLYGNGGNGFSQTTAGLTGGTGGSAGL
IGNGGNGGAGGAGANGGAGGNGGWLYGSGGNGGAGGAGPAGAIGAPGVAGGAGGAGGT
AGLFGNGGVGGVGGDGGQGGNGAGAGASGTKGGDAGAGGAGGAGGWIHGHGGAGGDGG
AGGAGGQASPGAPGPPSQPGGAGGAGGAGGRGGDGGSAGWLSGNGGDAGNGGGGGTAG
GAGNGGQFGGDGGTGGTGGTAGAGGNGGRGAVLFGHGGNAGHGGAGGNGAAAGAGGEH
VVATAGKGGTGGVGGDGGGGGAGGGGGLLYGNGGAGGAGNSGGDGGTGLNAALGGNGG
GGGVGGNAGAGGTGGSAGWLSGNGGAGGSGGSAGAGGAGGKGGDTPNGLAINPGIGGN
GGDTGNAGNGGNGGSAARLFGGGGAGGAGGTGSTAGSGGSGGTNPPTGLQAAGGNGGS
GHAGGHGGNGGGAGLLGGGGTGGNGGGGGQGGLGAAAGGVDGNGGNGGNGGKGGDAQL
VGDGGNGGNGGKGGAGLIAGLDGAGGAGGTRGLIFGNAGTPGQ"

misc_feature complement(16840. .16914)
/gene="PE_PGRS"
/note="PS00583 pfkB family of carbohydrate kinases signature 1"

RBS complement(18227. .18231)
/gene="PE_PGRS"
/note="possible RBS, gagga, for Rv1067c"

CDS complement(18555. .19946)
/gene="PE_PGRS"
/note="Rv1068c, (MTV017.21c), len: 463. Member of M. tuberculosis PE-family, Gly-, Ala-rich PGRS subfamily. FASTA scores: gp|Z80108|MTCY21B4_13 (576 aa) opt: 1689 z-score: 1003.3 E(): 0; 56.6% identity in 516 aa overlap. TBparse score is 0.849"
/codon_start=1
/transl_table=11
/product="PE_PGRS"
/protein_id="CAA17184.1"
/db_xref="PID:e1251947"
/db_xref="PID:g2896705"
/db_xref="GI:2896705"
/db_xref="SPTREMBL:O53416"
/translation="MSYMIAVPDMLSSAAGDLASIGSSINASTRAAAAATTRLLPAAA
DEVSAHIAALFSGHGEGYQAIARQMAAFHDQFTLALTSSAGAYASAEATNVEQQVLGL
INAPTQALLGRPLIGNGADGTAANPNGGAGGLLYGNGGNGFSQTTAGLTGGTGGSAGL
IGNGGNGGAGGAGANGGAGGNGGWLYGSGGNGGAGGAGPAGAIGAPGVAGGAGGAGGT
AGLFGNGGAGGAGGAGGAGGRGGDGGSAGWLSGNGGDAGTGGGGGNAGNGGNGGSAGW
LSGNGGTGGGGGTAGAGGQGGNGNSGIDPGNGGQGADTGNAGNGGHGGSAAKLFGDGG
AGGAGGMGSTGGTGGGGGFGGGTGGNGGNGHAGGAGGSGGTAGLLGSGGSGGTGGDGG
NGGLGAGSGAKGNGGNGGDGGKGGDAQLIGNGGNGGNGGKGGTGLMPGINGTGGAGGS
RGQISGNPGTPGQ"

RBS complement(19956. .19959)
/gene="PE_PGRS"
/note="possible RBS, ggag, for Rv1068c"

gene complement(20308. .22071)
/gene="Rv1069c"

CDS complement(20308. .22071)
/gene="Rv1069c"
/note="Rv1069c, (MTV017.22c), len: 587. Unknown, hydrophobic regions in N-terminal domain, but similar in part to Mycobacterium leprae B1306.04c protein. FASTA scores: sptr|O07136|O07136 B1306.04C PROTEIN (89 aa) opt: 229 z-score: 271.3 E(): 1.3e-07;54.2% identity in 72 aa

```
                overlap. TBparse score is 0.907"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv1069c"
                /protein_id="CAA17185.1"
                /db_xref="PID:e1251948"
                /db_xref="PID:g2896706"
                /db_xref="GI:2896706"
                /db_xref="SPTREMBL:O53417"
                /translation="MTEPAAATTTNASDEPATGAEQAVDTAATPQTPEPQPIRSTWWI
                RHYTFTGTAMGLVFVWFSMTPSLLPRGPLFQGLVSGICGAFGYGLGVFAVWLVRYMRS
                HNSSPPPPRWAWPPLIAVGAVGMVGMAVQFHVWQDDVRDLMGVEHLRWYDYPLAAALS
                LVVLFTLVEIGQFIRWLFRFLVGQVDRIAPFRVSAAIVVVLLVVLTITLLNGVVLKFA
                MNSMNSTFAAVNNEMNPDSAPPKTPLRSGGPGSLVSWESLGHQGRIFVHSGPTIADLT
                AFNGTPAVEPIRTYAGLNSADGIMATAELAARELARTGGLRRAVVAVATSTGTGWINE
                AEASALEYMYNGDTAIVSMQYSFLPSWLSFLVDKENARHAGEALFEAVDKLIRQLPES
                QRPKLVVFGESLGSFGGEAPFMNLNNILARTDGALFSGPTFNNTVWNSLTANRDAGSP
                QWLPIYDDGRNVRFVARARDLQRPDAPWGRPRVVYLQHASDPIAWWTPRLLFREPDWL
                REQRGYDVLPQTRWIPVVTFVQVSADMAVATHVPDGHGHRYVATVADGWAAVLSPPGW
                TQQKTERLQPLLHANAKPFGS"
gene            complement(22068. .22841)
                /gene="echA8"
CDS             complement(22068. .22841)
                /gene="echA8"
                /note="Rv1070c, (MTV017.23c), len: 257. echA8, Probable
                enoyl-CoA hydratase,equivalent to Mycobacterium leprae
                B1306.05c protein  (257aa), very similar to Rhizobium
                melilotii enoyl-CoA hydrataseL39265|RHMRPST_2  (257 aa).
                Contains enoyl-CoA hydratase. FASTA scores:
                sptr|O07137|O07137 B1306.05CPROTEIN (257 aa)opt: 1417
                z-score: 1737.8 E(): 0;   86.4% identity in 257 aaoverlap.
                TBparse score is 0.881. L39265|RHMRPST_2 Rhizobiummeliloti
                (257 aa) opt: 1100 z-score: 1351.2 E(): 0; 66.9%identity
                in 257 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="echA8"
                /protein_id="CAA17186.1"
                /db_xref="PID:e1251949"
                /db_xref="PID:g2896707"
                /db_xref="GI:2896707"
                /db_xref="SPTREMBL:O53418"
                /translation="MTYETILVERDQRVGIITLNRPQALNALNSQVMNEVTSAATELD
                DDPDIGAIIITGSAKAFAAGADIKEMADLTFADAFTADFFATWGKLAAVRTPTIAAVA
                GYALGGGCELAMMCDVLIAADTAKFGQPEIKLGVLPGMGGSQRLTRAIGKAKAMDLIL
                TGRTMDAAEAERSGLVSRVVPADDLLTEARATATTISQMSASAARMAKEAVNRAFESS
                LSEGLLYERRLFHSAFATEDQSEGMAAFIEKRAPQFTHR"
RBS             complement(22848. .22851)
                /gene="PE_PGRS"
                /note="possible RBS, gagg, for Rv1070c"
gene            complement(22853. .23890)
                /gene="echA9"
CDS             complement(22853. .23890)
                /gene="echA9"
                /note="Rv1071c, (MTV017.24c), len: 345. echA9, Probable
                enoyl-CoA hydratase or crotonase equivalent to
                Mycobacterium lepraecosmid B1306.06c protein (345 aa),
                similar to many eukaryotic and prokaryotic enoyl-CoA
                hydratases e.g. C. elegans Q19278 (386). FASTA scores:
                Y13803|MLB1306_6 Mycobacterium leprae (345 aa) opt: 1799
                z-score: 2130.4 E():  0;    77.7% identity in 345 aa
                overlap; and sptr|Q19278|Q19278 SIMILAR TOENOYL-COA
                HYDRATASES. (386 aa) opt:  787 z-score: 934.7 E(): 0;
                38.5% identity in 348 aa overlap. TBparse score is0.892"
                /codon_start=1
                /transl_table=11
                /product="echA9"
                /protein_id="CAA17187.1"
                /db_xref="PID:e1251950"
                /db_xref="PID:g2896708"
                /db_xref="GI:2896708"
                /db_xref="SPTREMBL:O53419"
                /translation="MTGESHEVLTNVEGGVGFVTLNRPKAINSLNQTMVDLLATVLMS
                WEHEDAVHAVVLSGAGERGLCAGGDVVAVYHSARKDGVEARRFWRHEYLLNALIGRFA
                KPYVALMDGIVMGGGVGVSAHANTRVVTDTSKVAMPEVGIGFIPDVGGVYLLSRAPGA
                LGLHAALTGAPFSGADAIALGFADHFVPHGDLDAFTQKIVTGGVESALAAHAVEPPPS
                TLAAQRDWIDECYAGDSVADIVAALRKQGGEPAVNASDLIASRSPIALSVTLQAVRRA
                AKLDTLEDVLIQDYRVSSASLRSHDLVEGIRAQLIDKDRNPNWSPATLDAITAADIEA
                YFEPVDDDLSF"
RBS             complement(23899. .23903)
                /gene="PE_PGRS"
                /note="possible RBS, gagga, for Rv1071c"
RBS             24065. .24069
                /gene="PE_PGRS"
                /note="possible RBS, gagga, for Rv1072"
gene            24077. .24913
                /gene="Rv1072"
CDS             24077. .24913
                /gene="Rv1072"
                /note="Rv1072, (MTV017.25), len: 278. Unknown
                transmembrane protein, equivalent to Mycobacterium leprae
                B1306.07 andU1740D protein fragments and similar to orf
                encoding PID:g1110441 (293aa) in bldA locus of
                Streptomyces griseus. . FASTAscores:sptr|O07139|O07139
                B1306.07 PROTEIN (FRAGMENT) (220aa) opt:1032 z-score:
                1240.4 E():  0;  75.0% identity in220 aa overlap; and
                M80628|STMBLDA_1 Streptomyces griseus (293 aa)opt:558
                z-score: 674.0 E(): 4.7e-30;    41.5%
                identityin299aaoverlap; and sptr|Q50056|Q50056 U1740D  (96
                aa)opt:381z-score: 469.6 E(): 1.2e-18; 71.6% identity in
                81 aaoverlap. TBparse score is 0.896"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv1072"
                /protein_id="CAA17188.1"
                /db_xref="PID:e1251951"
                /db_xref="PID:g2896709"
                /db_xref="GI:2896709"
                /db_xref="SPTREMBL:O53420"
                /translation="MRETSNPVFRSLPKQRGGYAQFGTGTAQQGFPADPYLAPYREAK
                ATRPLTIDDVVTKTGLTLAMLAGTAVVSYFLVASNVALAMPLTLVGALGGLALVLVAT
                FGRKQDNPAIVLSYAALEGLFLGAISFVLANFTVASANAGVLIGEAILGTMGVFFGML
                VVYKTGAIRVTPKFTRMVVAALFGVLVLMLGNLVLAMFNVGGGEGLGLRSPGPLGIIF
                SLVCIGIAAFSFLIDFDAADQMIRAGAPEKAAWGVALGLTVTLVWLYIEILRLLSYLQ
                NE"
RBS             25025. .25028
                /gene="PE_PGRS"
                /note="possible RBS, ggag, for Rv1073"
gene            25029. .25880
                /gene="Rv1073"
CDS             25029. .25880
                /gene="Rv1073"
                /note="Rv1073, (MTV017.26), len: 283. Unknown but
                similarto several hypothetical mycobacterial proteins
                including: Mycobacterium tuberculosis MTCY277.03 and
                MTCY06G11.02c (289aa); and Mycobacterium avium strain
                GIR10 AF002133_10 (346aa). FASTAscores: Z79701|MTCY277_3
                (339 aa) opt: 810 z-score: 1019.6 E():  0;    47.4%
                identity in 272 aa overlap;andAF0021|AF002133_10
                Mycobacterium avium strain GIR10 (346 aa) opt:802 z-score:
                1009.5 E():  0;    48.1% identity in 270aaoverlap; and
                Z92774|MTCY6G11_2  (289 aa) opt:  704z-score: 887.9 E():
                0;    44.4% identity in 259 aa overlap. TBparse score is
                0.942"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv1073"
                /protein_id="CAA17189.1"
                /db_xref="PID:e1251952"
                /db_xref="PID:g2896710"
                /db_xref="GI:2896710"
                /db_xref="SPTREMBL:O53421"
                /translation="MGAQPFIGSEALAAGLISWHELGKYYTAIMPNVYLDKRLKPSLR
                QRVIAAWLWSGRKGVIAGASASALHGAKWVDDHALVELIWRNARAPNGVRTKDELLLD
                GEVQRLCGLTVTTVERTAFDLGRRPPLGQAITRLDALANATDFKINDVRELARKHPHT
                RGLRQLDKALDLVDPGAQSPKETWLRLLLINAGFPRPSTQIPLLGVYGHPKYFLDMGW
```

EDIMLAVEYDGEQHRLSRDQFVKDVERLEYIRRAGWTHIRVLADHKGPDVVRRVRQAW
DTLTSRR"
gene complement(25954. .27171)
/gene="fadA3"
CDS complement(25954. .27171)
/gene="fadA3"
/note="Rv1074c, (MTV017.27c), len: 405. fadA3, Probable beta-ketoacyl CoA thiolase similar to many involved in beta-oxidation e.g. Escherichia coli X97452|ECPAA_12 (401 aa) and D90778_5 (401 aa), or Acinetobacter calcoaceticus ACCCAT_9 (401 aa). Contains PS00737 Thiolases signature 2 and PS00445 FGGY family of carbohydrate kinases signature 2, although this is probably fortuitous. FASTA scores: X97452|ECPAA_12 (EC 2.3.1.16) D90778_5 (401 aa) (Fatty oxidation complex betasubunit) D90778_5D90778_5(401 aa) opt: 1043 z-score:1198.1 E(): 0; 43.4% identity in 415 aa overlap; and M76991|ACCCAT_9 (401 aa) opt: 992 z-score: 1139.8 E(): 0; 41.2% identity in 415 aa overlap. TBparse score is 0.875"
/codon_start=1
/transl_table=11
/product="fadA3"
/protein_id="CAA17190.1"
/db_xref="PID:e1251953"
/db_xref="PID:g2896711"
/db_xref="GI:2896711"
/db_xref="SPTREMBL:O53422"
/translation="MPEAVIVSTARSPIGRAMKGSLVGMRPDDLAVQMVRAALDKVPA
LNPHQIDDLMMGCGLPGGESGFNIARVVAVALGYDFLPGTTVNRYCSSSLQTTRMAFH
AIKAGEGDAFISAGVETVSRFAKGNSDSWPDTKNPLFDGAQERSAAAAAGADEWHDPR
TDQKLPDIYIAMGQTAENVAIMTGISREEQDRWGVRSQNRAEEAIKNGFFEREITPVT
LPDGTTVSTDDGPRPGTTYEKVSELKPAFRPNGTVTAGNACPLNDGAAAVVITSDTKA
KELGLTPLARIVSTGVSGLSPEIMGLGPIEASKKALERAGMAITDIDLVEINEAFAVQ
VLGSARELGIDEDKLNISGGAIALGHPFGMTGARITTTLLNNLQTYDKTFGLETMCVG
GGQGMAMVIERLA"
misc_feature complement(26074. .26124)
/gene="fadA3"
/note="PS00737 Thiolases signature 2"
misc_feature complement(27052. .27114)
/gene="fadA3"
/note="PS00445 FGGY family of carbohydrate kinases signature 2"
RBS complement(27177. .27180)
/gene="PE_PGRS"
/note="possible RBS, ggag, for Rv1074c"
gene complement(27224. .28168)
/gene="Rv1075c"
CDS complement(27224. .28168)
/gene="Rv1075c"
/note="Rv1075c, (MTV017.28c), len: 314. Unknown but similar to Mycobacterium leprae U1740c (106 aa). Also weak similarity to many glyceraldehyde-3-phosphate dehydrogenases e.g.Q41595|G3PC_TAXBA (340 aa). FASTA scores: U15183|MLU15183_2Mycobacterium leprae cosmid B1740 (106 aa) opt: 207 z-score: 251.6 E(): 1.6e-06; 42.6% identity in 101 aa overlap;andsp|Q41595|G3PC_TAXBA GLYCERALDEHYDE 3-PHOSPHATE DEHYDRONENASE (340aa) opt: 147 z-score: 175.8 E(): 0.027; 27.5%identityin 189 aa overlap. TBparse score is 0.916"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1075c"
/protein_id="CAA17191.1"
/db_xref="PID:e1251954"
/db_xref="PID:g2896712"
/db_xref="GI:2896712"
/db_xref="SPTREMBL:O53423"
/translation="MPRRSTIALATAGALASTGTAYLGARNLLVGQATHARTVIPKSF
DAPPRADGVYTRGGGPVQRWRREVPFDVHLMIFGDSTATGYGCASAEEVPGVLIARGL
AEQTGKRIRLSTKAIVGATSKGVCGQVDAMFVVGPPPDAAVIMIGANDITALNGIGPS
AQRLADCVRRLRTRGAVVVVGTCPDLGVITAIPQPLRALAHTRGVRLARAQTAAVKAA
GGVPVPLGHLLAPKFRAMPELMFSADRYHPSAPAYALAADLLFLALRDALTEKLDIPI
HETPSRPGTATLEPGHTRHSMMSRLRRPRPARAVPTGG"
gene 28565. .29458
/gene="lipU"
CDS 28565. .29458
/gene="lipU"
/note="Rv1076, (MTV017.29), len: 297. Unknown but very similar to several Mycobacterium tuberculosis proteins e.g. MTCY13E12.41c (277 aa), and to esterases and lipases of around 300 aa e.g. Acinetobacter lwoffii esterase Q44087 (303).FASTA scores: Z95390|MTY13E12_40 (277 aa) opt: 1225 z-score: 1478.5 E(): 0; 76.0% identity in 246 aa overlap; andsptr|Q44087|Q44087 ESTERASE PRECURSOR (303 aa) opt: 427z-score: 519.5 E(): 1.9e-21; 32.5% identity in 280 aa overlap. TBparse score is 0.935"
/codon_start=1
/transl_table=11
/product="lipU"
/protein_id="CAA17192.1"
/db_xref="PID:e1251955"
/db_xref="PID:g2896713"
/db_xref="GI:2896713"
/db_xref="SPTREMBL:O53424"
/translation="MAVRPVLAVGSYLPHAPWPWGVIDQAARVLLPASTTVRAAVSLP
NASAQLVRASGVLPADGTRRAVLYLHGGAFLTCGANSHGRLVELLSKFADSPVLVVDY
RLIPKHSIGMALDDCHDGYRWLRLLGYEPEQIVLAGDSAGGYLALALAQRLQEVGEEP
AALVAISPLLQLAKEHKQAHPNIKTDAMFPARAFDALDALVASAAARNQVDGEPEELY
EPLEHITPGLPRTLIHVSGSEVLLHDAQLAAAKLAAAGVPAEVRVWPGQVHDFQVAAS
MLPEAIRSLRQIGEYIREATG"
gene 29515. .30909
/gene="cysM2"
CDS 29515. .30909
/gene="cysM2"
/note="Rv1077, (MTV017.30), cysM, len: 464. cysM2, Probable cystathionine beta-synthase similar throughout its lengthto many eukaryotic cystathionine beta-synthases, also similar inN-terminal domain (aa 1 - 330) to Mycobacterium tuberculosis CysK, MTCY98.03 (310 aa). Contains PS00217 Sugar transport proteinssignature 2 probably spurious. FASTA scores:sp|P32232|CBS_RAT CYSTATHIONINE BETA-SYNTHASE (EC 4.2. (560aa)opt: 951 z-score: 1063.6 E(): 0; 40.2% identity in450aa overlapgp|Z83860|MTCY98_3 (310 aa) opt: 855 z-score: 960.2 E(): 0; 46.8% identity in 314 aa overlap. TBparse score is 0.891"
/codon_start=1
/transl_table=11
/product="cysM2"
/protein_id="CAA17193.1"
/db_xref="PID:e1251956"
/db_xref="PID:g2896714"
/db_xref="GI:2896714"
/db_xref="SPTREMBL:O53425"
/translation="MRIAQHISELIGGTPLVRLNSVVPDGAGTVAAKVEYLNPGGSSK
DRIAVKMIEAAEASGQLKPGGTIVEPTSGNTGVGLALVAQRRGYKCVFVCPDKVSEDK
RNVLIAYGAEVVVCPTAVPPHDPASYYSVSDRLVRDIDGAWKPDQYANPEGPASHYVT
TGPEIWADTEGKVTHFVAGIGTGGTITGAGRYLKEVSGGRVRIVGADPEGSVYSGGAG
RPYLVEGVGEDFWPAAYDPSVPDEIIAVSDSDSFDMTRRLAREEAMLVGGSCGMAVVA
ALKVAEEAGPDALIVVLLPDGGRGYMSKIFNDAWMSSYGFLRSRLDGSTEQSTVGDVL
RRKSGALPALVHTHPSETVRDAIGILREYGVSQMPVVGAEPPVMAGEVAGSVSERELL
SAVFEGRAKLADAVSAHMSPPLRMIGAGELVSAAGKALRDWDALMVVEEGKPVGVITR
YDLLGFLSEGAGRR"
misc_feature 30043. .30120
/gene="cysM2"
/note="PS00217 Sugar transport proteins signature 2"
RBS 31098. .31102
/gene="PE_PGRS"
/note="possible RBS, ggaag, for Rv1078"
gene 31111. .31833
/gene="pra"
CDS 31111. .31833
/gene="pra"
/note="Rv1078, (MTV017.31), pra, len: 240. Unknown but equivalent to M.leprae proline rich antigen (249 aa). FASTA scores: X65546|MLPRAG_1 M.leprae pra gene for

proline rich antigen (249 aa) opt: 1162 z-score: 676.8 E(): 3.3e-30;64.8% identity in 253 aa overlap. TBparse score is 0.929"
/codon_start=1
/transl_table=11
/product="pra"
/protein_id="CAA17194.1"
/db_xref="PID:e1251957"
/db_xref="PID:g2896715"
/db_xref="GI:2896715"
/db_xref="SPTREMBL:O53426"
/translation="MTEQPPPGGSYPPPPPPPGPSGGHEPPPAAPPGGSGYAPPPPPS SGSGYPPPPPPPGGGAYPPPPPSAGGYAPPPPGPAIRTMPTESYTPWITRVLAAFIDW APYVVLVGIGWVIMLVTQTSSCVTSISEYDVGQFCVSQPSMIGQLVQWLLSVGGLAYL VWNYGYRQGTIGSSIGKSVLKFKVVSETTGQPIGFGMSVVRQLAHFIDAIICFVGFLF PLWDAKRQTLADKIMTTVCVPI"

gene 31865. .33031
/gene="metB"

CDS 31865. .33031
/gene="metB"
/note="Rv1079, (MTV017.32), metB, len: 388. Probable cystathionine gamma-synthase similar to many. closest to M. leprae P46807|METB_MYCLE (388 aa). Contains PS00868 Cys/Met metabolism enzymes pyridoxal-phosphate attachment site. FASTAscores: sp|P46807|METB_MYCLE CYSTATHIONINE GAMMA-SYNTHASE (EC 4.2.1.22) (388 aa) opt: 2220 z-score: 2621.3 E(): 0;87.3%identity in 387 aa overlap. TBparse score is 0.909"
/codon_start=1
/transl_table=11
/product="metB"
/protein_id="CAA17195.1"
/db_xref="PID:e1251958"
/db_xref="PID:g2896716"
/db_xref="GI:2896716"
/db_xref="SPTREMBL:O53427"
/translation="MSEDRTGHQGISGPATRAIHAGYRPDPATGAVNVPIYASSTFAQ DGVGGLRGGFEYARTGNPTRAALEASLAAVEEGAFARAFSSGMAATDCALRAMLRPGD HVVIPDDAYGGTFRLIDKVFTRWDVQYTPVRLADLDAVGAAITPRTRLIWVETPTNPL LSIADITAIAELGTDRSAKVLVDNTFASPALQQPLRLGADVVLHSTTKYIGGHSDVVG GALVTNDEELDEEFAFLQNGAGAVPGPFDAYLTMRGLKTLVLRMQRHSENACAVAEFL ADHPSVSSVLYPGLPSHPGHEIAARQMRGFGGMVSVRMRAGRRAAQDLCAKTRVFILA ESLGGVESLIEHPSAMTHASTAGSQLEVPDDLVRLSVGIEDIADLLGDLEQALG"

misc_feature 32462. .32506
/gene="metB"
/note="PS00868 Cys/Met metabolism enzymes pyridoxal-phosphate attachment site"

gene complement(33102. .33596)
/gene="greA"

CDS complement(33102. .33596)
/gene="greA"
/note="Rv1080c, (MTV017.33c), greA, len: 164. Probable transcription elongation factor G similar to many e.g P21346|GREA_ECOLI (158 aa) closest to M. leprae P46808|GREA_MYCLE (202 aa). Contains two PS00829 and one PS00830 Prokaryotic transcriptionelongation factors signatures 1 and 2, respectively. FASTAscores: sp|P46808|GREA_MYCLE TRANSCRIPTION ELONGATION FACTORG (202 aa) opt: 1005 z-score: 1178.3 E(): 0;94.5% identity in 164 aa overlap; and sp|P21346|GREA_ECOLITRANSCRIPTION ELONGATION FACTOR G (158 aa) opt: 257 z-score: 313.4 E(): 5.7e-10; 37.2% identity in 148 aa overlap. TBparse score is 0.892"
/codon_start=1
/transl_table=11
/product="greA"
/protein_id="CAA17196.1"
/db_xref="PID:e1251959"
/db_xref="PID:g2896717"
/db_xref="GI:2896717"
/db_xref="SPTREMBL:O53428"
/translation="MTDTQVTWLTQESHDRLKAELDQLIANRPVIAAEINDRREEGDL RENGGYHAAREEQGQQEARIRQLQDLLSNAKVGEAPKQSGVALPGSVVKVYYNGDKSD SETFLIATRQEGVSDGKLEVYSPNSPLGGALIDAKVGETRSYTVPNGSTVSVTLVSAE PYHS"

misc_feature complement(33177. .33227)
/gene="greA"
/note="PS00830 Prokaryotic transcription elongation factors signature 2"

misc_feature complement(33447. .33569)
/gene="greA"
/note="PS00829 Prokaryotic transcription elongation factors signature 1"

RBS complement(33606. .33610)
/gene="PE_PGRS"
/note="possible RBS, aggag, for Rv1080c"

gene complement(33782. .34216)
/gene="Rv1081c"

CDS complement(33782. .34216)
/gene="Rv1081c"
/note="Rv1081c, (MTV017.34c), len: 144. Unknown but hydrophobic stretch from aa 26 - 48. TBparse score is 0.906"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1081c"
/protein_id="CAA17197.1"
/db_xref="PID:e1251960"
/db_xref="PID:g2896718"
/db_xref="GI:2896718"
/db_xref="SPTREMBL:O53429"
/translation="MTHTPIPRPDARYGRPRLSRRARRRVAIALGVLVAAAGIVIAVI GYQRISTSAVTGSLVGYRLVDDETASVTISVTRSDPSRPVACIVRVRATNGSETGRRE LLVPPSEATTVQVTTTVKSSQPPVMADVYGCGTEVPSYLRLP"

RBS 34308. .34314
/gene="PE_PGRS"
/note="possible RBS, aagggg, for Rv1082"

gene 34318. .35184
/gene="Rv1082"

CDS 34318. .35184
/gene="Rv1082"
/note="Rv1082, (MTV017.35), len: 288. Unknown but equivalent to Mycobacterium leprae LmbE (290 aa) similar to Mycobacterium tuberculosis MTV005.06 (303 aa) MTV005.06, and similar to Streptomyces lincolnensis lmbE gene TR:Q54358 (EMBL:X79146) from Lincomycin production gene cluster (270 aa).FASTA scores: gp|U15183|MLU15183_8 Mycobacterium leprae cosmid B1740 (290 aa) opt: 1737 z-score: 2046.5 E(): 0; 86.4% identity in 287 aa overlap; and AL0101|MTV005_6 (303 aa)opt: 411 z-score: 489.2 E(): 9.4e-20; 35.8% identity in299aa overlap. TBparse score is 0.896"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1082"
/protein_id="CAA17198.1"
/db_xref="PID:e1251961"
/db_xref="PID:g2896719"
/db_xref="GI:2896719"
/db_xref="SPTREMBL:O53430"
/translation="MSELRLMAVHAHPDDESSKGAATLARYADEGHRVLVVTLTGGER GEILNPAMDLPDVHGRIAEIRRDEMTKAAEILGVEHTWLGFVDSGLPKGDLPPPLPDD CFARVPLEVSTEALVRVVREFRPHVMTTYDENGGYPHPDHIRCHQVSVAAYEAAGDFC RFPDAGEPWTVSKLYYVHGFLRERMQMLQDEFARHGQRGPFEQWLAYWDPDHDFLTSR VTTRVECSKYFSQRDDALRAHATQIDPNAEFFAAPLAWQERLWPTEEFELARSRIPAR PPETELFAGIEP"

gene 35181. .35447
/gene="Rv1083"

CDS 35181. .35447
/gene="Rv1083"
/note="Rv1083, (MTV017.36), len: 88. Unknown but similar to hypothetical Mycobacterium leprae protein U1740E (167 aa). FASTA scores: U15183|MLU15183_9 Mycobacterium leprae cosmid B1740 (167 aa) opt: 332 z-score: 379.2 E(): 1.2e-13; 58.4% identity in 101 aa overlap. TBparse score is 0.905"

```
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv1083"
                /protein_id="CAA17199.1"
                /db_xref="PID:e1251962"
                /db_xref="PID:g2896720"
                /db_xref="GI:2896720"
                /db_xref="SPTREMBL:O53431"
                /translation="MNQILLSVIAEGGPGNTGPDFGKASPVGLLVIVLLVIATLFLVR
                SMNQQLKKVPKSFDRDHPELDQAADEGTDRDGPARPPGPPHESG"
gene            35434. .37455
                /gene="Rv1084"
CDS             35434. .37455
                /gene="Rv1084"
                /note="Rv1084, (MTV017.37), len: 673. Unknown but
                similarto conserved hypothetical proteins from B.
                subtilis, P37512|YYAL_BACSU (689 aa), Archaeoglobus
                fulgidus AE000983_10 (642 aa), etc. FASTA scores:
                sp|P37512|YYAL_BACSU HYPOTHETICAL78.8 KD PROTEIN IN T (689
                aa)  opt: 1063 z-score: 1253.9E(): 0;36.5% identity in 696
                aa overlap; and gp|AE0009|AE000983_10 Archaeoglobus
                fulgidus section 1 (642 aa) opt: 1018 z-score: 1201.1 E():
                0;    37.2% identity in 600 aa overlapTBparsescore is
                0.872"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv1084"
                /protein_id="CAA17200.1"
                /db_xref="PID:e1251963"
                /db_xref="PID:g2896721"
                /db_xref="GI:2896721"
                /db_xref="SPTREMBL:O53432"
                /translation="MSPANPSGTNTLALATSPYLRQHADNPVHWQQWTPQALAEAAAR
                AVPILLSVGYAACHWCHVMAHESFDDDEVAAAMNAGFVCIKVDREERPDIDAVYMNAT
                VALTGQGGWPMTCFLTPNGRPFFCGTYYPKAAFLQLLSAISETWRERRAEVEQASDHI
                AAELRSMASGLPGGGPEVAPELCDDAVAGVLREQDTAHGGFGGAPKFPPSALLEALMR
                HYERTRSPAALEAVARTGNAMARGGIYDQLGGGFARYSVDGAWVVPHFEKMLYDNALL
                LRAYAHWARRTGDPLARRVAAQTARFLLDELGSKAPADMFTSSLDADADGREGSTYVW
                TPVQLTEVLGGDDGRWAAEVFGVTEAGTFEHGTSVLQLPADPDDAARLDRVRAALLVA
                RLARAQPARDDKVVTSWNGLAITALAEASVALDDPALAHAARRCATRLLDLHVVDGRL
                RRASLGGVVGDSAAILEDHAMLATGLLALYQLTSEGAWLTAATGLLDTAVAHFGDPQR
                PGRWFDTADDAERLMLRPSDPLDGATPSGASSIAEALLTAGHVVDGARAERYWQLAAD
                TLRAHAVLLARAPRSAGHWLAVAEAVVRGPLQIAVACDLPRSSLLADARRLAPGGAIV
                VGGAAGSSALLVGRDRVAGADAAYVCRGRVCDLPVTSAAELATALGVPG"
gene            complement(37554. .38282)
                /gene="Rv1085c"
CDS             complement(37554. .38282)
                /gene="Rv1085c"
                /note="Rv1085c, (MTV017.38c), len: 242. Unknown but
                similar to hypothetical proteins from Escherichia coli
                ECU28375_49 (219 aa),Borrelia burgdorferi HIAE001124_2
                (233 aa), etc.FASTA scores: gp|U28375|ECU28375_49  (219
                aa) opt:  308 z-score: 401.0 E(): 7.5e-15;    30.6%
                identity in 180 aa overlap; and AE0011|HIAE001124_2
                Borrelia burgdorferi (233 aa)opt:  305 z-score: 396.9 E():
                1.3e-14;    25.6% identity in203 aa overlap. TBparse score
                is 0.884. Also weakly similartohaemolysin from Bacillun
                spp. FASTA scores; HLY3_BACCEP54176 haemolysin iii.
                bacillus cereus. (219 aa), opt: 247z-score: 341.0 E():
                8.7e-12, 27.5% identity in 171 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv1085c"
                /protein_id="CAA17201.1"
                /db_xref="PID:e1251964"
                /db_xref="PID:g2896722"
                /db_xref="GI:2896722"
                /db_xref="SPTREMBL:O53433"
                /translation="MSGQADTATTAEARTPAHAAHHLVEGVARVLTKPRFRGWIHVYS
                AGTAVLAGASLVAVSWAVGSAKAGLTTLAYTAATITMFTVSATYHRVNWKSATARNWM
                KRADHSMIFVFIAGSYTPFALLALPAHDGRVVLSIVWGGAIAGILLKMCWPAAPRSVG
                VPLYLLLGWVAVWYTATILHNAGVTALVLLFVGGALYSIGGILYAVRWPDPWPTTFGY
                HEFFHACTAVAAICHYIAMWFVVF"

gene            38393. .39181
                /gene="Rv1086"
CDS             38393. .39181
                /gene="Rv1086"
                /note="Rv1086, (MTV017.39), len: 262. Unknown but member
                of the UPF0015 FAMILY similar to family members from
                Streptomyces fradiae P20182|YT14_STRFR (259 aa) and
                Corynebacterium glutamicicum P38118|YARF_CORGL (234 aa).
                Contains PS01066 Uncharacterized protein family UPF0015
                signature. FASTA scores: sp|P20182|YT14_STRFR HYPOTHETICAL
                29.1 KD PROTEIN (259 aa)  opt:  840 z-score: 1078.8 E():
                0; 51.0% identityin 259 aaoverlap; and
                sp|P38118|YARF_CORGL HYPOTHETICAL 26.1 KD PROTEIN (234 aa)
                opt:  729 z-score: 937.8 E(): 0; 56.0% identity in 209 aa
                overlap. Also similar to Rv2361c(MTCY27.19)(296 aa,35.6%
                identity in 233 aa overlap)"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv1086"
                /protein_id="CAA17202.1"
                /db_xref="PID:e1251965"
                /db_xref="PID:g2896723"
                /db_xref="GI:2896723"
                /db_xref="SPTREMBL:O53434"
                /translation="MEIIPPRLKEPLYRLYELRLRQGLAASKSDLPRHIAVLCDGNRR
                WARSAGYDDVSYGYRMGAAKIAEMLRWCHEAGIELATVYLLSTENLQRDPDELAALIE
                IITDVVEEICAPANHWSVRTVGDLGLIGEEPARRLRGAVESTPEVASFHVNVAVGYGG
                RREIVDAVRALLSKELANGATAEELVDAVTVEGISENLYTSGQPDPDLVIRTSGEQRL
                SGFLLWQSAYSEMWFTEAHWPAFRHVDFLRALRDYSARHRSYGR"
misc_feature    39011. .39067
                /gene="Rv1086"
                /note="PS01066 Uncharacterized protein family UPF0015
                signature"
RBS             39347. .39350
                /gene="PE_PGRS"
                /note="possible RBS, ggag, for Rv1087"
CDS             39358. .41661
                /gene="PE_PGRS"
                /note="Rv1087, (MTV017.40), len: 767. Member of M.
                tuberculosis PE family, Gly-, Ala-rich PGRS subfamily.
                Contains PS00583 pfkBfamily of carbohydrate kinases
                signature 1 near C-terminus.FASTA scores:
                sp|P71933|YQ04_MYCTU HYPOTHETICAL 63.1 KD GLYCINE-RICH
                (778 aa)  opt: 2365 z-score: 1413.2 E(): 0;    53.5%
                identity in 809 aa overlap. TBparse score is0.859"
                /codon_start=1
                /transl_table=11
                /product="PE_PGRS"
                /protein_id="CAA17203.1"
                /db_xref="PID:e1251966"
                /db_xref="PID:g2896724"
                /db_xref="GI:2896724"
                /db_xref="SPTREMBL:O53435"
                /translation="MSFVVVAPEVLAAAASDLAGIGSTLAQANAAALAPTTAVLAAGA
                DEVSAAIASLFGAHGQAYQAVSAQMSAFHAQFMQALTGAGGAYAAAEAVNVSAAQSVE
                QDLLAAINARFERIFGRPLIGDGANGGPGQDGGPGGLLYGNGGNGGTSTTVGMAGGNG
                GAAGLIGNGGFGGGGGPGAAGGNGGAGGWLFGNGGAGGAGGLGVAPGVPGGAGGAGGA
                GGVGGPAGLWGHGGAGGAGGAGVAGAGGFEGTIGAGGAGGVGGAGGVGGAGGAGGWLY
                GDAGAGGDGGVGGAGGTGGLGNRGGAGGAGGAGGVGGAGGAAGLWGGGGAGGVGGTGG
                GAGLGAQSVTFSSSLSGLSGGDGGAGGAGGAGGAGGTGGWLYGGGGAAGSGGDGGTGG
                QGGAGGAGVFSLFGSGGGPGGNGGVGGVGGVGGAGGRAGLFGVGGLGGAGGDAGDSGE
                GGFGGPGLAGGLFGNPGNGGVGGIGGDAAAGGAGGAGGNGGAGGNGGWLFGNGGAGGS
                GGDGGAAGRGGAGNLGSAGGINAPAGNPGSGSVGIGGAGGAGGTAGLFGDGGAGGAGG
                AGAAGGFGGISAATPSAGSEGAMGGAGGVGGNARLLGTGGAGGVGGGGGAGGDGGRGG
                VATPGGQGGDAGDGGAGGAGGNGGGASGAGGWLLGTGGAGGAGGNGGNGGKAGFSPGP
                TNFGLNGAGGGGVGGNGATGPWLFGDGGPTPGSTGAGAAGGHGGDAQLIGNGGHGGA
                GGTGVPNGSGGAGGLSGLLFGEPGANG"
misc_feature    41437. .41511
                /gene="PE_PGRS"
                /note="PS00583 pfkB family of carbohydrate kinases
                signature 1"
RBS             42300. .42304
                /gene="PE_PGRS"
```

/note="possible RBS, aagga, for Rv1088"
gene 42311. .42929
/gene="PE"
CDS 42311. .42745
/gene="PE"
/note="Rv1088, (MTV017.41), len: 144. Member of Mycobacterium tuberculosis PE family, similar to many others e.g. MTCI418B_6 (487 aa) except it appears to be frameshifted around codon 84. No error to account for frameshift could be found. FASTA scores: gp|Z96071|MTCI418B_6 Mycobacterium tuberculosis cosmid (487 aa) opt: 318 z-score: 383.5 E(): 7.3e-14; 60.9% identity in 87 aa overlap. TBparse score is 0.943"
/codon_start=1
/transl_table=11
/product="PE"
/protein_id="CAA17204.1"
/db_xref="PID:e1251967"
/db_xref="PID:g2896725"
/db_xref="GI:2896725"
/db_xref="SPTREMBL:O53436"
/translation="MSYMIATPAALTAAATDIDGIGSAVSVANAAAVAATTGVLAAGG DEVLAAIARLFNANAEEYHALSAQVAAFQTLFVRTLTGGCGVFRRRRGRQCVTAAEHR AAGAGRRQRRRRSGDGQWRLRQQRHFGCGGQPEFRQHSEHRR"
CDS 42567. .42929
/gene="PE"
/note="Protein sequence is in conflict with the conceptual translation; Rv1089, (MTV017.42), len: 120. Partial orf that appears to be frameshifted continuation of MTV017.41. Sequencehas been checked and appears correct. FASTA scores: Z95555|MTCY06F7_4Mycobacterium tuberculosis cosmid (401 aa) opt:126 z-score: 142.1 E(): 2; 29.6% identity in 125 aa overlap"
/codon_start=1
/transl_table=11
/product="PE"
/protein_id="CAA17205.1"
/db_xref="PID:e1251968"
/db_xref="PID:g2896726"
/db_xref="GI:2896726"
/db_xref="SPTREMBL:O53437"
/translation="SFAGAEAANASQLQSIARQVRGAVNAVAGQVTGNGGSGNSGTSA AAANPNSDNTASIADRGTSAIMTTASATASSTGVDGGIAATYAVASQWDGGYVANYTI TQFGRDFDDRLAVAIHFA"
misc_feature 43243. .43383
/gene="PE_PGRS"
/note="region showing similarity to part of coding sequence for cellulase B precursor TR:Q54331. Maybe vestige of cellulase gene that was once contiguous with next orf, MTV017.43."
gene 43397. .43852
/gene="Rv1090"
CDS 43397. .43852
/gene="Rv1090"
/note="Rv1090, (MTV017.43), len: 151. Probable cellulase or endoglucanase fragment similar to many e.g. O08468 CEL2 (EC 3.2.1.4) (CELLULASE) ENDO (377 aa). Gene appears to havebeen inactivated by frameshift mutations but no errors could be found that would account for this. FASTA scores: sptr|O08468|O08468 CEL2 (EC 3.2.1.4) (CELLULASE) (ENDO (377 aa)opt: 554z-score: 684.7 E(): 1.2e-30; 52.0% identity in152 aa overlap. TBparse score is 0.876"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1090"
/protein_id="CAA17206.1"
/db_xref="PID:e1251969"
/db_xref="PID:g2896727"
/db_xref="GI:2896727"
/db_xref="SPTREMBL:O53438"
/translation="MGTNLPTEVGQILSAPTSIDYNYPTTGVWDASYDICLDSTPKTT GVNQQEIMIWFNHQGSIQPVGSPVGNTTIEGKNFVVWDGSNGMNNAMAYVATEPIEVW SFDVMSFVDHTATMEPITDSWYLTSIRAGLEPWSDGVGLGVDSFSAKVN"

RBS 44256. .44260
/gene="PE_PGRS"
/note="possible RBS, ggagg, for Rv1091"
CDS 44267. .46828
/gene="PE_PGRS"
/note="Rv1091, (MTV017.44), len: 853. Member of M. tuberculosis PE-family, Gly-, Ala-rich PGRS subfamily. FASTA scores: sptr|O06810|O06810 HYPOTHETICAL 107.4 KD PROTEIN. (1329aa) opt:2723 z-score: 1545.5 E(): 0; 52.6% identity in951 aa overlap. TBparse score is 0.859"
/codon_start=1
/transl_table=11
/product="PE_PGRS"
/protein_id="CAA17207.1"
/db_xref="PID:e1251970"
/db_xref="PID:g2896728"
/db_xref="GI:2896728"
/db_xref="SPTREMBL:O53439"
/translation="MSFVIAAPEALVAVASDLAGIGSALAEANAAALAPTTALLAAGA DEVSAAIAALFGAHGQAYQTVSAQASAFHAQFVQALTGGGAYAAAEAANVSAAQSTD QRLLDLINGPTQALLGRPLIGDGANGGPGQDGGPGGLLYGNGGNGGTSTTAGVAGGNG GAAGLIGNGGAGGGGGAGAAGGNGGAGGWLYGNGGAGGAGGTSVIPGVAGGNGGAGGS AGLWGTGGAGGDGGNGRSGPVNVAGSAGGNGGAGGAAGLFGDAGAGGNGGKGGAGGAA FSINFTAGDGGAGGAGGSGGHALLWGAGGAGGNGGSGGTGGAGGSTAGAGGNGGAGGG GGTGGLLFGNGGAGGHGAAAGNGLAAGNGVSSSGGGGAGGTGGAGGDGGAGGAGGNAR LWGVGGAGGAGGDGGAGGAGGKGGSGLSGNANGGAGGDSGRGGTGGAGGEGGAAGLLV GTGGHGGDGGAGGAAVKGGDGGAAAGTGIAGAGGRGGAGGSGGSGGDGGGAAGPAGW LFGDGGAGGNGGAAAAGGAGGQAGGGGGNGGNGGNGGNGGNGGNGATGGWLYGNGGAG GQGATAGAGGAGANGVSSTNGGGTGGNGGIGGTGGSGGAGGNAGLLGVGGAGGHGASG GAGDRGGAGGTGFISSDGGAGGDGGDGGNGGAGGTGGLLFGAGGNGGPGGSGGAADIG GNGGAGNGGGTDGNGGNGGSGGGAGSGGDGGGAGGNGAWLFGNGGAGGGGGKGGNGAG GGLGGGSFGLPGLNGSGGDGGDGGNGAPGGVLYGNGGAGGQGSSGGIGGPGATGGAGG KGGDGGDAQLIGDGGNGGNGGAGGTGGTPGPGGPGGSGGLGGLLFGQTGTAGVSP"
gene complement(47046. .47984)
/gene="coaA"
CDS complement(47046. .47984)
/gene="coaA"
/note="Rv1092c, (MTV017.45c), coaA, len: 312. Probable pantothenate kinase similar to many e.g. P15044|COAA_ECOLI (316 aa). Contains PS00017 ATP/GTP-binding site motif A (P-loop). FASTA scores: sp|P15044|COAA_ECOLI PANTOTHENATE KINASE(EC 2.7.1.33) (316 aa) opt: 1079 z-score: 1353.6 E(): 0; 52.7% identity in 311 aa overlap. TBparse score is 0.912"
/codon_start=1
/transl_table=11
/product="coaA"
/protein_id="CAA17208.1"
/db_xref="PID:e1251971"
/db_xref="PID:g2896729"
/db_xref="GI:2896729"
/db_xref="SPTREMBL:O53440"
/translation="MSRLSEPSPYVEFDRRQWRALRMSTPLALTEEELVGLRGLGEQI DLLEVEEVYLPLARLIHLQVAARQRLFAATAEFLGEPQQNPDRPVPFIIGVAGSVAVG KSTTARVLQALLARWDHHPRVDLVTTDGFLYPNAELQRRNLMHRKGFPESYNRRALMR FVTSVKSGSDYACAPVYSHLHYDIIPGAEQVVRHPDILILEGLNVLQTGPTLMVSDLF DFSLYVDARIEDIEQWYVSRFLAMRTTAFADPESHFHHYAAFSDSQAVVAAREIWRTI NRPNLVENILPTRPRATLVLRKDADHSINRLRLRKL"
misc_feature complement(47673. .47696)
/gene="coaA"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
gene 48372. .49652
/gene="glyA"
CDS 48372. .49652
/gene="glyA"
/note="Rv1093, (MTV017.46), glyA, len: 426. Probable SERINE HYDROXYMETHYLTRANSFERASE (EC 2.1.2.1) similar to many, best with HYPHOMICROBIUM METHYLOVORUM P34895|GLYA_HYPME (434aa). FASTAscores: sp|P34895|GLYA_HYPME SERINE HYDROXYMETHYLTRANSFERASE (434 aa) opt: 1492 z-score: 1757.8 E(): 0; 56.8% identity in 419 aa overlap. TBparse score is 0.880"
/codon_start=1

```
                /transl_table=11
                /product="glyA"
                /protein_id="CAA17209.1"
                /db_xref="PID:e1251972"
                /db_xref="PID:g2896730"
                /db_xref="GI:2896730"
                /db_xref="SPTREMBL:O53441"
                /translation="MSAPLAEVDPDIAELLAKELGRQRDTLEMIASENFVPRAVLQAQ
                GSVLTNKYAEGLPGRRYYGGCEHVDVVENLARDRAKALFGAEFANVQPHSGAQANAAV
                LHALMSPGERLLGLDLANGGHLTHGMRLNFSGKLYENGFYGVDPATHLIDMDAVRATA
                LEFRPKVIIAGWSAYPRVLDFAAFRSIADEVGAKLLVDMAHFAGLVAAGLHPSPVPHA
                DVVSTTVHKTLGGGRSGLIVGKQQYAKAINSAVFPGQQGGPLMHVIAGKAVALKIAAT
                PEFADRQRRTLSGARIIADRLMAPDVAKAGVSVVSGGTDVHLVLVDLRDSPLDGQAAE
                DLLHEVGITVNRNAVPNDPRPPPMVTSGLRIGTPALATRGFGDTEFTEVADIIATALAT
                GSSVDVSALKDRATRLARAFPLYDGLEEWSLVGR"
gene            49757. .50584
                /gene="desA2"
CDS             49757. .50584
                /gene="desA2"
                /note="Rv1094, (MTV017.47), len: 275. desA2, Probable
                stearoyl-acyl carrier protein desaturase weakly similar to
                plant stearoyl-acyl carrier protein desaturases but very
                similar to Mycobacterium tuberculosis enzyme DES (338 aa)
                and to Mycobacterium leprae  AadX (338 aa). FASTA scores:
                U49839|MTU49839_1 Mycobacterium tuberculosis DES (d (338
                aa) opt:  525 z-score: 669.4E(): 8.5e-30;    32.2%
                identity in 270 aaoverlap; and U15182|MLU15182_32  (338
                aa) opt: 506 z-score: 645.4 E(): 1.9e-28;    34.1%
                identity in 261 aa overlap.TBparse score is0.894"
                /codon_start=1
                /transl_table=11
                /product="desA2"
                /protein_id="CAA17210.1"
                /db_xref="PID:e1251973"
                /db_xref="PID:g2896731"
                /db_xref="GI:2896731"
                /db_xref="SPTREMBL:O53442"
                /translation="MAQKPVADALTLELEPVVEANMTRHLDTEDIWFAHDYVPFDQGE
                NFAFLGGRDWDPSQSTLPRTITDACEILLILKDNLAGHHRELVEHFILEDWWGRWLGR
                WTAEEHLHAIALREYLVVTREVDPVANEDVRVQHVMKGYRAEKYTQVETLVYMAFYER
                CGAVFCRNLAAQIEEPILAGLIDRIARDEVRHEEFFANLVTHCLDYTRDETIAAIAAR
                AADLDVLGADIEAYRDKLQNVADAGIFGKPQLRQLISDRITAWGLAGEPSLKQFVTG"
RBS             50784. .50788
                /note="possible RBS, aggag, for Rv1095"
gene            50795. .52096
                /gene="phoH2"
CDS             50795. .52096
                /gene="phoH2"
                /note="Rv1095, (MTV017.48), len: 433. Unknown but
                similarto many proteins described as PhoH-like,
                e.g.Bacillus subtilis BSZ97025_12 (442 aa) or M.
                tuberculosis O05830|PHOL_MYCTU (352 aa). Contains PS00017
                ATP/GTP-binding site motif A(P-loop).FASTA scores:
                gp|Z97025|BSZ97025_12 (442 aa) opt: 605 z-score: 737.5
                E(): 0;    40.1% identity in 444 aa overlap;
                andsp|O05830|PHOL_MYCTU PHOH-LIKE PROTEIN.  (352 aa)opt:
                390 z-score: 477.8 E(): 4e-19;    31.5% identity in241aa
                overlap. TBparse score is 0.885"
                /codon_start=1
                /transl_table=11
                /product="phoH2"
                /protein_id="CAA17211.1"
                /db_xref="PID:e1251974"
                /db_xref="PID:g2896732"
                /db_xref="GI:2896732"
                /db_xref="SPTREMBL:O53443"
                /translation="MTDTRTYVLDTSVLLSDPWACSRFAEHDVVVPLVVISELEAKRH
                HHELGWFARQALRLFDDLRLEHGRLDQPIPVGTQGGTLHVELNHTDPAVLPAGFRTDS
                NDSRILSCAANLAAEGKRVTLVSKDIPLRVKAAAVGLAADEYHAQDVVVSGWSGMHEL
                ETASADIDALFADGEIDLVEARDLPCHTGIRLLGGGSHALGRVNAHKRVQLVRGDREA
                FGLRGRSAEQRVALDLLLDESVGIVSLGGKAGTGKSALALCAGLEAVLERRTHRKVVV
                FRPLYAVGGQELGYLPGSESEKMGPWAQAVFDTLEGLASPAVLEEVLSRGMLEVLPLT
                HIRGRSLHDSFVIVDEAQSLERNVLLTVLSRLGTGSRVVLTHDIAQRDNLRVGRHDGV
                AAVIEKLKGHPLFAHITLLRSERSPIAALVTEMLEEITGPR"
misc_feature    51533. .51556
                /gene="phoH2"
                /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
misc_feature    52106. .52199
                /note="Conserved intergenic region, similar to positions
                (+1) 34333 34426 of MTCY427, and (+1) 6762 6824 MTCI376"
gene            52183. .53058
                /gene="Rv1096"
CDS             52183. .53058
                /gene="Rv1096"
                /note="Rv1096, (MTV017.49), len: 291. Unknown but
                similarto parts of diverse enzymes involved in
                carbohydrate degradationiincluding Bacillus subtilis
                protein similar to endo-1,4-beta-xylanase (467 aa),
                Streptomyces lividans acetyl-xylan esterase (334 aa), etc.
                FASTA scores: gp|Z99110|BSUB0007_92  (467aa) opt: 418
                z-score: 445.1 E(): 2.6e-17;    38.6% identity in 184 aa
                overlap and gp|M64552|STMXLNB_2 Streptomyces lividans
                acetyl-xylan (334 aa) opt:  371 z-score: 398.1
                E():1.1e-14;    31.6% identity in 237 aa overlap. TBparse
                score is 0.905"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv1096"
                /protein_id="CAA17212.1"
                /db_xref="PID:e1251975"
                /db_xref="PID:g2896733"
                /db_xref="GI:2896733"
                /db_xref="SPTREMBL:O53444"
                /translation="MPKRPDNQTWRYWRTVTGVVVAGAVLVVGGLSGRVTRAENLSCS
                VIKCVALTFDDGPGPYTDRLLHILTDNDAKATFFLIGNKVAANPAGARRIADAGMEIG
                SHTWEHPNMTTIPPEDIPGQFSRANDVIAAATGRTPTLYRPAGGLSNDAVRQAAAKVG
                QAEILWDVIPFDWINDSNTAATRHMLMTQIKPGSVVLFHDTYSSTVDVVYQFIPVLKA
                NGYRLVTVSELLGPRAPGSSYGSRENGPPVNELRDIPASEIPPLPNTSSPKPMPNFPI
                TDIAGQNSGGPNNGA"
gene            complement(53061. .53942)
                /gene="Rv1097c"
CDS             complement(53061. .53942)
                /gene="Rv1097c"
                /note="Rv1097c, (MTV017.50c), len: 293. Gly-, Pro-rich
                protein similar to Mycobacterium tuberculosis MTCY07A7.13
                andlike MTCY07A7.13  contains potential membrane spanning
                region (aa ~68 - 92). FASTA scores: Z95556|MTCY07A7_13
                Mycobacterium tuberculosis cosmi (273 aa)  opt:  219
                z-score: 176.8E():  0.023;    30.5% identity in 266 aa
                overlap.  TBparsescore is 0.912"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv1097c"
                /protein_id="CAA17213.1"
                /db_xref="PID:e1251976"
                /db_xref="PID:g2896734"
                /db_xref="GI:2896734"
                /db_xref="SPTREMBL:O53445"
                /translation="MTVPPAGPYGNYPYGPNTYGQDPYWGGQPQGGSYPPAYPPQQYP
                PGWPAGPYPPGPPPPGPGSKTPWLILAGLAVLGVILLVVILVIGLRGDNKSTTATSPA
                TSAPTSQPFSQQTATGCTPNVSGGVQPIGDSISAGKLSFPTSAAPGWSAFSDDQNPNL
                IDAVGVGHEVAGADQWMMQAEVAITNFVTTMDVAAQASKLMQCVADGPGYAGSSPTLG
                PTKTSSITVDGVRAARVDADITIADSSRNVKGDSVTIIAVDTKPVTVFLGATPIGDAT
                SRATVERVIEALKVNKS"
gene            complement(53939. .55363)
                /gene="fum"
CDS             complement(53939. .55363)
                /gene="fum"
                /note="Rv1098c, (MTV017.51c), fum, len: 474. Probable
                fumarase, similar to many P14408|FUMH_RA (507 aa) or
                P05042|FUMC_ECOLI (467 aa). FASTA scores:
                sp|P14408|FUMH_RAT FUMARATEHYDRATASE, MITOCHONDRIAL
                PRECURSOR (507 aa)  opt: 1427 z-score: 1659.8 E(): 0;
                52.3% identity in 461 aa overlap;andsp|P05042|FUMC_ECOLI
                FUMARATE HYDRATASE CLASS II (467 aa)opt: 1355 z-score:
                1576.7 E(): 0;    50.2% identity in 444aaoverlap. Contains
```

PS00163 Fumarate lyases signature.TBparse score is 0.886"
/codon_start=1
/transl_table=11
/product="fum"
/protein_id="CAA17214.1"
/db_xref="PID:e1251977"
/db_xref="PID:g2896735"
/db_xref="GI:2896735"
/db_xref="SPTREMBL:O53446"
/translation="MAVDADSANYRIEHDTMGEVRVPAKALWRAQTQRAVENFPISGR
GLERTQIRALGLLKGACAQVNSDLGLLAPEKADAIIAAAAEIADGQHDDQFPIDVFQT
GSGTSSNMNTNEVIASIAAKGGVTLHPNDDVNMSQSSNDTFPTATHIAATEAAVAHLI
PALQQLHDALAAKALDWHTVVKSGRTHLMDAVPVTLGQEFSGYARQIEAGIERVRACL
PRLGELAIGGTAVGTGLNAPDDFGVRVVAVLVAQTGLSELRTAANSFEAQAARDGLVE
ASGALRTIAVSLTKIANDIRWMGSGPLTGLAEIQLPDLQPGSSIMPGKVNPVLPEAVT
QVAAQVIGNDAAIAWGGANGAFELNVYIPMMARNILESFKLLTNVSRLFAQRCIAGLT
ANVEHLRRLAESSPSIVTPLNSAIGYEEAAAVAKQALKERKTIRQTVIDRGLIGDRLS
IEDLDRRLDVLAMAKAEQLDSDRL"

misc_feature complement(54386. .54415)
/gene="fum"
/note="PS00163 Fumarate lyases signature"

RBS complement(55373. .55377)
/note="possible RBS, aagga, for Rv1098c"

gene complement(55394. .56380)
/gene="Rv1099c"

CDS complement(55394. .56380)
/gene="Rv1099c"
/note="Rv1099c, (MTV017.52c), len: 328. Unknown but highlysimilar to conserved protein called GlpX (believed to be involved in glycerol metabolism e.g. Q03224|YWJI_BACSU (321aa) and P44811|GLPX_HAEIN (333 aa). Start chosen on basis ofsimilarity, others possible upstream e.g. ATG at 57050 chosen by TBparse. FASTA scores: sp|Q03224|YWJI_BACSU HYPOTHETICAL 34.0 KD PROTEIN IN R (321 aa) opt: 1092 z-score: 1303.2 E(): 0; 52.1% identity in 313 aa overlap; and sp|P44811|GLPX_HAEIN GLPX PROTEIN HOMOLOG (333 aa) opt: 763 z-score: 912.1E():0; 46.2% identity in 327 aa overlap. TBparse scoreis0.878"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1099c"
/protein_id="CAA17215.1"
/db_xref="PID:e1251978"
/db_xref="PID:g2896736"
/db_xref="GI:2896736"
/db_xref="SPTREMBL:O53447"
/translation="MELVRVTEAGAMAAGRWVGRGDKEGGDGAAVDAMRELVNSVSMR
GVVVIGEGEKDHAPMLYNGEEVGNGDGPECDFAVDPIDGTTLMSKGMTNAISVLAVAD
RGTMFDPSAVFYMNKIAVGPDAAHVLDITAPISENIRAVAKVKDLSVRDMTVCILDRP
RHAQLIHDVRATGARIRLITDGDVAGAISACRPHSGTDLLAGIGGTPEGIIAAAAIRC
MGGAIQAQLAPRDDAERRKALEAGYDLNQVLTTEDLVSGENVFFCATGVTDGDLLKGV
RYYPGGCTTHSIVMRSKSGTVRMIEAYHRLSKLNEYSAIDFTGDSSAVYPLP"

gene 56481. .57182
/gene="Rv1100"

CDS 56481. .57182
/gene="Rv1100"
/note="Rv1100, (MTV017.53), len: 233. Unknown but slightlysimilar to Mycobacterium tuberculosis hypothetical proteinMTCY180.12 (156 aa). Start uncertain. FASTA scores: Z97193|MTCY180_12 Mycobacterium tuberculosis cosmid (156 aa) opt: 122 z-score: 132.5 E(): 6.9; 27.4% identity in 135 aaoverlap. TBparse score is 0.914"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1100"
/protein_id="CAA17216.1"
/db_xref="PID:e1251979"
/db_xref="PID:g2896737"
/db_xref="GI:2896737"
/db_xref="SPTREMBL:O53448"
/translation="MVGDCPRSRTVRWSWDTGHVTAEPQPTPRPAKPRLLQDGRDMFW
SLAPLVVGCILLAGLVGMCSFQLGGTKRGPIPSYDAAQALRADAKTLGFPIRLPQLPG
GWTPNSGGRGGIENGRADPATGQRRNAATSIVGFISPTGRYLSLTQSNADEDKLVGSI
HPSMYPTGTVDVGGTRWVVYEGSDENGAVEPVWTTRLTGPGGATQLAITGAGSIDQFR
TLASATQSQPPLPAR"

gene complement(57189. .58346)
/gene="Rv1101c"

CDS complement(57189. .58346)
/gene="Rv1101c"
/note="Rv1101c, (MTV017.54c), len: 385. Probable membraneprotein shows some similarity to other bacterial proteins e.g. P77406|PERM_ECOLI (353 aa). FASTA scores: sp|P77406|PERM_ECOLI PUTATIVE PERMEASE PERM (353 aa) opt: 287 z-score:346.0 E(): 8.8e-12; 24.9% identity in 349 aa overlap.TBparse score is 0.886"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1101c"
/protein_id="CAA17217.1"
/db_xref="PID:e1251980"
/db_xref="PID:g2896738"
/db_xref="GI:2896738"
/db_xref="SPTREMBL:O53449"
/translation="MNTEFTLTQKRALAILTLIALLFGAYFLRNYFVLIVVAAVGAYL
FTPLFKWFTKRFNTGLSAACTLLSALAAVVVPVGALVGLAIVQIARMVDSVADWVRTT
DLSTLGDKILQFVNGLFDRVPFLHITVTADALRKAMISVAQNVGEWLLHFLRDAAGSL
AGVITSAIIFVYVFVALLVNREKLRTLIGQLNPLGEDVTDLYLQKMGSMVRGTVNGQF
VIAACQGVAGAASIYIAGFHHGFFIFAIVLTALSIIPLGGGIVTIPFGIGMIFYGNIA
GGIFVLLWHLLVVTNIDNVLRPILVPRDARLNSALMLLSVFAGITMFGPWGIIIGPVL
MILIVTTIDVYLAVYKGVELEQFEAPPVRRRWLPRRGPATSRNAPPPSTAE"

RBS complement(58353. .58358)
/note="possible RBS, ggcgga, for Rv1101c"

gene complement(58458. .58769)
/gene="Rv1102c"

CDS complement(58458. .58769)
/gene="Rv1102c"
/note="Rv1102c, (MTV017.55c), len: 103. Unknown but similar to Mycobacterium tuberculosis hypothetical protens: MTI376.17 (102aa) and MTCY09F9.22 (109 aa). FASTA scores: Z84498|MTCY9F9_22(109 aa) opt: 158 z-score: 227.3 E(): 3.6e-05; 33.3%identity in 93 aa overlap and Z95972|MTCI376_17 (102aa) opt: 140 z-score: 203.9 E(): 0.00072; 40.6% identity in69aa overlap. TBparse score is 0.924"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1102c"
/protein_id="CAA17218.1"
/db_xref="PID:e1251981"
/db_xref="PID:g2896739"
/db_xref="GI:2896739"
/db_xref="SPTREMBL:O53450"
/translation="MRPIHIAQLDKARPVLILTREVVRPHLTNVTVAPITTTVRGLAT
EVPVDAVNGLNQPSVVSCDNTQTIPVCDLGRQIGYLLASQEPALAEAIGNAFDLDWVV
A"

gene complement(58769. .59089)
/gene="Rv1103c"

CDS complement(58769. .59089)
/gene="Rv1103c"
/note="Rv1103c, (MTV017.56c), len: 106. Unknown, TBparse score is 0.916"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1103c"
/protein_id="CAA17219.1"
/db_xref="PID:e1251982"
/db_xref="PID:g2896740"
/db_xref="GI:2896740"
/db_xref="SPTREMBL:O53451"
/translation="MYLPWGVVLAGGANGFGAGAYQTGTICEVSTQIAVRLPDEIVAF
IDDEVRGQHARSRAAVVLRALERERRRRLAERDAEILATNTSATGDLDTLAGHCARTA
LDID"

CDS 59099. .59788
/gene="Rv1104"
/note="Rv1104, (MTV017.57), len: 229. Possible esterase similar to the N-terminal domain of many e.g. B. subtilis

```
                P37967|PNBA_BACSU (489 aa). Gene may be inactivated as
                similarity continues in other frames. FASTA scores:
                sp|P37967|PNBA_BACSU PARA-NITROBENZYL ESTERASE (EC 3.1
                (489 aa) opt:  715 z-score:790.7 E():  0;    53.4%
                identity in 191 aa overlap. TBparsescore is 0.934"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv1104"
                /protein_id="CAA17220.1"
                /db_xref="PID:e1251983"
                /db_xref="PID:g2896741"
                /db_xref="GI:2896741"
                /db_xref="SPTREMBL:O53452"
                /translation="MVVDSCVAESRYGPVRGADDGRVKVWKGIRYAAPPLGDLRFRTP
                EPPERWTEVADATTFGPACPQPAIPNMPLDLGASQSEDCWSLNIWAPADTEPGDGKPV
                MVWLHGGAYILGSGSQPLYNGRRLAASGDVVVVTVNYRLGALGFLDLSSFNTSRRRFD
                SNIGLRDVLAVLRWVADNIAVFGGDPEKVTLFGESARESSRPCSPPRRPRVCSRRRSP
                RAHRRHRSTTR"
misc_feature    59099. .60621
                /note="region corresponding to B. subtilis gene encoding
                sptr|P71048|P71048 PARA-NITROBENZYL ESTERASE (489 aa).
                Several frameshifts appear to have occurred. Sequence has
                been checked but no errors found."
gene            59099. .59788
                /gene="Rv1104"
gene            60082. .60624
                /gene="Rv1105"
CDS             60082. .60624
                /gene="Rv1105"
                /note="Protein sequence is in conflict with the conceptual
                translation; Rv1105, (MTV017.58), len: 180. Partial orf
                similar to C-terminal domain of P71048 PARA-NITROBENZYL
                ESTERASE (489 aa). Probably inactivated gene as several
                frameshifts required to obtain product continuing from
                MTV017.57. FASTA scores: sptr|P71048|P71048
                PARA-NITROBENZYL ESTERASE (489 aa) opt:  248 z-score:
                319.4 E(): 2.7e-10; 32.3% identity in167 aa overlap.
                TBparse score is 0.936"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv1105"
                /protein_id="CAA17221.1"
                /db_xref="PID:e1251984"
                /db_xref="PID:g2896742"
                /db_xref="GI:2896742"
                /db_xref="SPTREMBL:O53453"
                /translation="CRSPRAITSMFTQIAAEQPDLQVPTEEQIGSAYSRWRRKARSLS
                MATDVGFRMPSVWLAEGHSGVAPVYLYRFDYSTPLLKLLLVRAAHATELPYVWGNLGG
                SQDPALKLGDAKAAIAVSRRVRTRWINFATRGKPTGPDGEPDWPCYEEAHRACLIIGR
                RDAVVHDVDAHIRATWGSKW"
gene            complement(60642. .61754)
                /gene="Rv1106c"
CDS             complement(60642. .61754)
                /gene="Rv1106c"
                /note="Rv1106c, (MTV017.59c), len: 370. Probable
                cholesterol dehydrogenase highly similar to Nocardia sp.
                NAD(P)-dependent cholesterol dehydrogenase Q03704 (364
                aa). Similarity suggest start at ATG at 62310 but then no
                RBS. Also similar to viral3-beta hydroxy-5-ene steroid
                dehydrogenase MCU32426_1 (354aa). FASTA scores:
                sptr|Q03704|Q03704 NAD(P)-DEPENDENT CHOLESTEROL DEHYDR
                (364 aa) opt: 1789 z-score: 2215.9E(): 0;  74.5% identity
                in 361 aa overlap; and gp|U32426|MCU32426_1 Molluscum
                contagiosum virus (354 aa) opt:  432 z-score: 538.2 E():
                1.7e-22; 34.6% identity in 347 aa overlap. TBparse score
                is 0.885"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv1106c"
                /protein_id="CAA17222.1"
                /db_xref="PID:e1251985"
                /db_xref="PID:g2896743"
                /db_xref="GI:2896743"
                /db_xref="SPTREMBL:O53454"
                /translation="MLRRMGDASLTTELGRVLVTGGAGFVGANLVTTLLDRGHWVRSF
                DRAPSLLPAHPQLEVLQGDITDADVCAAAVDGIDTIFHTAAIIELMGGASVTDEYRQR
                SFAVNVGGTENLLHAGQRAGVQRFVYTSSNSVVMGGQNIAGGDETLPYTDRFNDLYTE
                TKVVAERFVLAQNGVDGMLTCAIRPSGIWGNGDQTMFRKLFESVLKGHVKVLVGRKSA
                RLDNSYVHNLIHGFILAAAHLVPDGTAPGQAYFINDAEPINMFEFARPVLEACGQRWP
                KMRISGPAVRWVMTGWQRLHFRFGFPAPLLEPLAVERLYLDNYFSIAKARRDLGYEPL
                FTTQQALTECLPYYVSLFEQMKNEARAEKTAATVKP"
RBS             complement(61757. .61760)
                /note="possible RBS, aagg, for Rv1106c"
gene            complement(61764. .62021)
                /gene="xseB"
CDS             complement(61764. .62021)
                /gene="xseB"
                /note="Rv1107c, (MTV017.60c), xseB, len: 85. Probable
                exonuclease VII small subunit similar to those from H.
                influenzae (P43914|EX7S_HAEIN 84 aa) and E. coli
                (P22938|EX7S_ECOLI79 aa). FASTA scores:
                sp|P43914|EX7S_HAEIN EXODEOXYRIBONUCLEASE SMALLSUBUNI (84
                aa) opt:  126 z-score: 187.5 E():  0.006;    37.3%
                identity in 67 aa overlap; and sp|P22938|EX7S_ECOLI
                EXODEOXYRIBONUCLEASE SMALL SUBUNI (79 aa) opt:
                125z-score: 186.6 E(): 0.0067;    39.7% identity in 58 aa
                overlap. TBparsescore is 0.912"
                /codon_start=1
                /transl_table=11
                /product="xseB"
                /protein_id="CAA17223.1"
                /db_xref="PID:e1251986"
                /db_xref="PID:g2896744"
                /db_xref="GI:2896744"
                /db_xref="SPTREMBL:O53455"
                /translation="MVCDPNGDDTGRTHATVPVSQLGYEACRDELMEVVRLLEQGGLD
                LDASLRLWERGEQLAKRCEEHLAGARQRVSDVLAGDEAQNG"
gene            complement(62011. .63258)
                /gene="xseA"
CDS             complement(62011. .63258)
                /gene="xseA"
                /note="Rv1108c, (MTV017.61c), xseA, len: 415. Probable
                exonuclease VII large subunit similar to those from E.
                coli (P04994|EX7L_ECOLI 456 aa), B. subtilis, H.
                influenzae and H.pylori. FASTA scores:
                sp|P04994|EX7L_ECOLI EXODEOXYRIBONUCLEASE LARGESUBUNI (456
                aa) opt:  581 z-score: 682.6 E(): 1.6e-30;30.8%identity in
                425 aa overlap. TBparse score is 0.890"
                /codon_start=1
                /transl_table=11
                /product="xseA"
                /protein_id="CAA17224.1"
                /db_xref="PID:e1251987"
                /db_xref="PID:g2896745"
                /db_xref="GI:2896745"
                /db_xref="SPTREMBL:O53456"
                /translation="MTQNSAENPFPVRAVAIRVAGWIDKLGAVWVEGQLAQITMRPDA
                KTVFMVLRDPAADMSLTVTCSRDLVLSAPVKLAEGVQVVVCGKPSFYTGRGTFSLRLS
                EIRAVGIGELLARIDRLRRLLDAEGLFDPRLKRPIPYLPNMIGLITGRASAAERDVTT
                VASARWPAARFAVRNVAVQGPNAVGQIVEALRELDRDPDVDVIVLARGGGSVEDLLPF
                SDETLCRAIAACRTPVVSAVGHEPDNPLCDLVVDLRAATPTDAAKKVVPDTAAEQRLI
                DDLRRRSAQALRNWVSREQRAVAQLRSRPVLADPMTMVSVRAEEVHRARSTLRRNLTL
                MVAAETERIGHLAARLATLGPAATLARGYAIVQTVAQTGPEGGSEPQVLRSVHDAPEG
                TKLRVRVADGALAAVSEGQTNGL"
RBS             complement(62029. .62033)
                /gene="xseA"
                /note="possible RBS, aagga, for Rv1107c"
gene            complement(63255. .63893)
                /gene="Rv1109c"
CDS             complement(63255. .63893)
                /gene="Rv1109c"
                /note="Rv1109c, (MTV017.62c), len: 212. Unknown. TBparse
                score is 0.905"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv1109c"
```

```
                     /protein_id="CAA17225.1"
                     /db_xref="PID:e1251988"
                     /db_xref="PID:g2896746"
                     /db_xref="GI:2896746"
                     /db_xref="SPTREMBL:O53457"
                     /translation="MATAPYGVRLLVGAATVAVEETMKLPRTILMYPMTLASQAAHVV
                     MRFQQGLAELVIKGDNTLETLFPPKDEKPEWATFDEDLPDALEGTSIPLLGLSDASEA
                     KNDDRRSDGRFALYSVSDTPETTTASRSADRSTNPKTAKHPKSAAKPTVPTPAVAAEL
                     DYPALTLAQLRARLHTLDVPELEALLAYEQATKARAPFQTLLANRITRATAK"
gene                 63983. .64990
                     /gene="lytB'"
CDS                  63983. .64990
                     /gene="lytB'"
                     /note="Rv1110, (MTV017.63), len: 335. Unknown but very
                     similar to LytB proteins from many bacteria (appears to
                     have N-terminal extension) e.g. sp|P22565|LYTB_ECOLI LYTB
                     PROTEIN(316 aa). Also very similar to another LytB-like
                     protein from M. tuberculosis MTV004.40c (329 aa). FASTA
                     scores: sp|P22565|LYTB_ECOLI LYTB PROTEIN (316 aa) opt:
                     1041 z-score: 1275.0 E():0;    52.4% identity in 309 aa
                     overlap; and AL0091|MTV004_38 (329 aa) opt:  975 z-score:
                     1194.2 E(): 0;51.3%identity in 312 aa overlap. TBparse
                     score is 0.899"
                     /codon_start=1
                     /transl_table=11
                     /product="lytB'"
                     /protein_id="CAA17226.1"
                     /db_xref="PID:e1251989"
                     /db_xref="PID:g2896747"
                     /db_xref="GI:2896747"
                     /db_xref="SPTREMBL:O53458"
                     /translation="MVPTVDMGIPGASVSSRSVADRPNRKRVLLAEPRGYCAGVDRAV
                     ETVERALQKHGPPVYVRHEIVHNRHVVDTLAKAGAVFVEETEQVPEGAIVVFSAHGVA
                     PTVHVSASERNLQVIDATCPLVTKVHNEARRFARDDYDILLIGHEGHEEVVGTAGEAP
                     DHVQLVDGVDAVDQVTVRDEDKVVWLSQTTLSVDETMEIVGRLRRRFPKLQDPPSDDI
                     CYATQNRQVAVKAMAPECELVIVVGSRNSSNSVRLVEVALGAGARAAHLVDWADDIDS
                     AWLDGVTTVGVTSGASVPEVLVRGVLERLAECGYDIVQPVTTANETLVFALPRELRSP
                     R"
gene                 complement(65007. .65921)
                     /gene="Rv1111c"
CDS                  complement(65007. .65921)
                     /gene="Rv1111c"
                     /note="Rv1111c, (MTV017.64c), len: 304. Unknown,
                     N-terminal domain hydrophobic, C-terminal half very rich
                     in Arg. TBparse score is 0.934"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein Rv1111c"
                     /protein_id="CAA17227.1"
                     /db_xref="PID:e1251990"
                     /db_xref="PID:g2896748"
                     /db_xref="GI:2896748"
                     /db_xref="SPTREMBL:O86351"
                     /translation="MPWWAAILIAVTATAIGYAIDAGSGHKALTLVFTGCYIAGCVGA
                     VLAVRQSDLFTALVQPPLILFCAVPGAYWLFHGGTIGKFKDLLINCGYSLIERFPLML
                     GTAAGVLLIGLVRWYLGTALFDSIARKLSSLMTGDSDDDGGRRSAQRPARTRSRHARP
                     PSEDNREPIAERRSRRRPRPQNDPHPRRNAHERPAPRSSRFDSYRSYQPSEPSGPAEP
                     VNRYERRGARYQPYARYEPTYEPQRRRARPSEPTNPTHHPISQVRYRGSATRDARRDN
                     YREEQRFDRRDRSRAPRRPPAESWEYDV"
gene                 66053. .67126
                     /gene="Rv1112"
CDS                  66053. .67126
                     /gene="Rv1112"
                     /note="Rv1112, (MTCY22G8.01-MTV017.65), len: 357.
                     Probablegtp binding protein, similar to eg. YCHF_HAEIN
                     P44681 probable gtp-binding protein (362 aa), fasta
                     scores; opt: 1189z-score: 1347.0 E(): 0 52.7% identity in
                     357 aa overlap, contains PS00017 ATP/GTP-binding site
                     motif A (P-loop)"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein Rv1112"
                     /protein_id="CAA17228.1"
                     /db_xref="PID:e1301417"
                     /db_xref="PID:g3256023"
                     /db_xref="GI:3256023"
                     /db_xref="SPTREMBL:O53459"
                     /translation="MSLSLGIVGLPNVGKSTLFNALTRNNVVAANYPFATIEPNEGVV
                     SLPDPRLDKLAELFGSQRVVPAPVTFVDIAGLVKGASEGAGLGNKFLAHIRECDAICQ
                     VVRVFVDDDVTHVTGRVDPQSDIEVVETELILADLQTLERATGRLEKEARTNKARKPV
                     YDAALRAQQVLDAGKTLFAAGVDAAALRELNLLTTKPFLYVFNADEAVLTDPARVGEL
                     RALVAPADAVFLDAAIESELTELDDESAAELLESIGQSERGLDALARAGFHTLKLQTF
                     LTAGPKEARAWTIHQGDTAPKAAGVIHSDFEKGFIKAEIVSYDDLVAAGSMAAAKAAG
                     KVRIEGKDYVMADGDVVEFRFNV"
     source          66070. .>67200
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37Rv"
                     /db_xref="taxon:1773"
                     /clone="Y22G8"
     misc_feature    66077. .66100
                     /gene="Rv1112"
                     /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
BASE COUNT    10714 a  21541 c  23321 g  11624 t
ORIGIN


  Query Match             3.8%;  Score 37.6;  DB 1;  Length 67200;
  Best Local Similarity   49.0%;  Pred. No. 12;
  Matches  100;  Conservative    0;  Mismatches  104;  Indels    0;  Gaps    0;

Qy      40 GACCCAGTGGATATCGGTGGCCCCGTTATTCGTCCAGGTGCCCAGGGAGGAGGACCCGCC 99
           | | | |  |  | |||||||  || |  |      || | |  ||  |  || ||||||
Db   45725 GGCGCGGGCGGCAGCGGTGGCAGCGGTGGTGACGGCGGGGGCGGGGCCGCCGGCCCCGCC 45784

Qy     100 TGCAGCATGAACCTGTGGCTCCTGGCCTGCCTGGTGGCCGGCTTCCTGGGAGCCTGGGCC 159
           |  |  ||  |        |  |||  ||  |   | |||   |  ||  ||| | | |
Db   45785 GGGTGGCTGTTCGGCGATGGCGGGGCTGGCGGGAACGGCGGGGCCGCGGCCGCCGGCGGC 45844

Qy     160 CCCGCTGTCCACGCCCAAGGTGTCTTTGAGGACTGCTGCCTGGCCTACCACTACCCCATT 219
            |||  | ||| |||   ||||  |   | | || || ||   | |  | ||  |  || |
Db   45845 GCCGGCGGCCAAGCCGGCGGTGGCGGCGGGAACGGCGGCAATGGCGGCAACGGCGGCAAT 45904

Qy     220 GGGTGGGCTGTGCTCCGGCGCGCC 243
           ||  |   ||    |    |||||
Db   45905 GGCGGCAATGGCGGCAACGGCGCC 45928


RESULT  11
SC2A11
LOCUS       SC2A11      22789 bp    DNA             BCT       05-AUG-1998
DEFINITION  Streptomyces coelicolor cosmid 2A11.
ACCESSION   AL031184
NID         g3402233
VERSION     AL031184.1  GI:3402233
KEYWORDS    ABC transporter; aminomethyltransferase; gcvH; gcvT; glyA; glycine
            cleavage system H protein; glycine cleavage system T protein;
            L-serine dehydratase; muramoyl-pentapeptide carboxypeptidase;
            oligopeptide transport system; pyridoxal-phosphate-dependent
            aminotransferase; sdaA; serine hydroxymethyltransferase;
            transmembrane transport protein; trmU; tRNA
            (5-methylaminomethyl-2-thiouridylate)-methyltransferase.
SOURCE      Streptomyces coelicolor.
  ORGANISM  Streptomyces coelicolor
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 22789)
  AUTHORS   Murphy,L. and Harris,D.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 22789)
  AUTHORS   Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-AUG-1998) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
            David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
            Colney, Norwich, Norfolk NR4 7UH, UK
```

```
REFERENCE   3  (bases 1 to 22789)
  AUTHORS   Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
            Kinashi,H. and Hopwood,D.A.
  TITLE     A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL   Mol. Microbiol. 21 (1), 77-96 (1996)
  MEDLINE   97000351
COMMENT     Notes:
            Streptomyces coelicolor sequencing at The Sanger Centre is funded
            by the BBSRC.
            Details of S. coelicolor sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL; http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are
            numbered using the following system eg SC7B7.01c. SC (S.
            coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
            strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous. The length
            in codons is given for each CDS.
            Usually the highest scoring match found by fasta -o is given for
            CDS which show significant similarity to other CDS in the database.
            The position of possible ribosome binding site sequences are given
            where these have been used to deduce the initiation codon. Gene
            prediction is based on positional base preference in codons using a
            specially developed Hidden Markov Model (Krogh et al., Nucleic
            Acids Research, 22(22):4768-4778(1994)) and the FramePlot program
            of Bibb et al., Gene 30:157-66(1984) as implemented at
            http://www.nih.go.jp/
            jun/cgi-bin/frameplot.pl. CAUTION:  We may not have predicted the
            correct initiation codon.  Where possible we choose an initiation
            codon (atg, gtg, ttg or (att)) which is preceded by an upstream
            ribosome binding site sequence (optimally 5-13bp before the
            initiation codon).  If this cannot be identified we choose the most
            upstream initiation codon.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone.  It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions. Cosmid 2A11 lies
            between 3D11 and 8D9 on the AseI-B genomic restriction fragment.
FEATURES             Location/Qualifiers
     source          1. .22789
                     /organism="Streptomyces coelicolor"
                     /strain="A3(2)"
                     /db_xref="taxon:1902"
                     /clone="cosmid 2A11"
     gene            1. .571
                     /gene="SC2A11.01"
     CDS             <1. .571
                     /gene="SC2A11.01"
                     /note="SC2A11.01, probable muramoyl-pentapeptide
                     carboxypeptidase, partial CDS, len: 189 aa; highly similar
                     to S. albus CBPM_STRAL muramoyl-pentapeptide
                     carboxypeptidase precursor (EC 3.4.17.8) (255 aa), fasta
                     scores; opt: 955 z-score: 1605.0 E(): 0, 71.1% identity in
                     190 aa overlap"
                     /codon_start=2
                     /transl_table=11
                     /product="muramoyl-pentapeptide carboxypeptidase"
                     /protein_id="CAA20170.1"
                     /db_xref="PID:e1314373"
                     /db_xref="PID:g3402234"
                     /db_xref="GI:3402234"
                     /db_xref="SPTREMBL:O86562"
                     /translation="IRVAGWVTSGERLSYDGQYGARTAAAVKKFQADYGLGADGVAGP
                     ATFSKIYALQDADCTPVHFTYAELNKCNSDWSGGAVSAATAKSNALKTMWKLEAMRHA
                     LGDVPITISSGFRSRACNSAVGGSSTSRHLYGDAADLTGSPSFCRLAQQARTHGFSEI
                     LGPGYPGHNDHTHVAFDPSPYWSAPNCGI"
     gene            complement(626. .2044)
                     /gene="SC2A11.02c"
     CDS             complement(626. .2044)
                     /gene="SC2A11.02c"
                     /note="SC2A11.02c, probable transmembrane transport
                     protein, len: 472 aa; similar to many e.g. YHJE_ECOLI
                     hypothetical metabolite transport protein (440 aa), fasta
                     scores; opt: 772 z-score: 1638.5 E(): 0, 56.7% identity in
                     441 aa overlap and SHIA_ECOLI shikimate transporter (438
                     aa), fasta scores; opt: 553 z-score: 980.0 E(): 0, 34.6%
                     identity in 437 aa overlap. Contains PS00216 Sugar
                     transport proteins signature 1 and Pfam match to entry
                     PF00083 sugar_tr, Sugar (and other) transporters, score
                     183.10, E-value 4.4e-51"
                     /codon_start=1
                     /transl_table=11
                     /product="transmembrane transport protein"
                     /protein_id="CAA20171.1"
                     /db_xref="PID:e1314374"
                     /db_xref="PID:g3402235"
                     /db_xref="GI:3402235"
                     /db_xref="SPTREMBL:O86563"
                     /translation="MPAPASVPAPQTPVNPRSRVLVASLMGTTIEFYDFYIYATAAVL
                     VFPKLFFPSSDPTTALLSSFAVFGAAMVARPIGAVFFGHLGDRLGRKKTLVVSLLTMG
                     IATFLIGALPTYAQAGWVATALLVLMRLAQGFALGGEWSGAALVATENAPRGKRALWG
                     TFPQLGAPLGFIIGNGLFLIIGALLPSESGADPTQPSEAFANWGWRIPFLFSAVMVAI
                     GLWVRSRLVESTVFTRTREAGKVRKLPLATVVQGHWKQLVLGTFIMLATYVLFYLMTT
                     FSLSYGRAAKDADVPGLGYSYTTFVLMMIFGVLFFAVFTLVSGPLADRYGRRATLVWI
                     TVAIVVFGLLWVPLIDMGTLGVVLWLVLGFTLMGMTFGPMGALLPELFPTSVRYTGSG
                     ISYNVSSILGAAVAPFIAVALWEAGDGSPWLVGVYLSAMAVLTLAALLLSKETKDVSL
                     EEHGSAPAGEDPRTAAATSSVS"
     misc_feature    complement(674. .1990)
                     /gene="SC2A11.02c"
                     /note="Pfam match to entry PF00083 sugar_tr, Sugar (and
                     other) transporters, score 183.10, E-value 4.4e-51"
     misc_feature    complement(1040. .1093)
                     /gene="SC2A11.02c"
                     /note="PS00216 Sugar transport proteins signature 1"
     gene            complement(2161. .3528)
                     /gene="sdaA"
     CDS             complement(2161. .3528)
                     /gene="sdaA"
                     /note="SC2A11.03c, sdaA, probable L-serine dehydratase,
                     len: 455 aa; highly similar to e.g. SDHL_ECOLI L-serine
                     dehydratase 1 (EC 4.2.1.13) (454 aa), fasta scores; opt:
                     1643 z-score: 1576.6 E(): 0, 55.6% identity in 459 aa
                     overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="L-serine dehydratase"
                     /protein_id="CAA20172.1"
                     /db_xref="PID:e1314375"
                     /db_xref="PID:g3402236"
                     /db_xref="GI:3402236"
                     /db_xref="SPTREMBL:O86564"
                     /translation="MAISVFDLFSIGIGPSSSHTVGPMRAARMFARRLRNEELLDSVA
                     SVRVELYGSLGATGHGHGTPKAVLLGLEGDSPRTVDVESADDRVETIKSSGRISLLGD
                     HEIAFAYDDDMVLHRRKALPYHANGMTLWAYDAEGAEVLTKTYYSVGGGFVVDEDAVG
                     ADRIVLDDTVLKYPFRTGDELLRLARETGLSISALMLENERAWRDEDEIREGLLEIWR
                     VMRACVDRGMTREGILPGGLKVRRRAANTARKLRSEGDPQALAMEWITLYAMAVNEEN
                     AAGGRVVTAPTNGAAGIIPAVLHYYMNFVPGADEDGVVRFLLAAGAIGMLFKENASIS
                     GAEVGCQGEVGSACSMAAGALAEVLGGSPEQVENAAEIGMEHNLGLTCDPVGGLVQIP
                     CIERNGMAAVKAVTAARMAMRGDGSHKVSLDKVIKTMKETGADMSVKYKETARGGLAV
                     NIIEC"
     RBS             complement(3534. .3538)
                     /note="possible RBS upstream of sdaA"
     misc_feature    3599. .3610
                     /note="single clone coverage only"
     gene            complement(3630. .4892)
                     /gene="glyA"
     CDS             complement(3630. .4892)
                     /gene="glyA"
                     /note="SC2A11.04c, glyA, probable serine
                     hydroxymethyltransferase, len: 420 aa; highly similar  to
                     many e.g. GLYA_ECOLI serine hydroxymethyltransferase (EC
                     2.1.2.1) (417 aa), fasta scores; opt: 1530 z-score: 1748.7
                     E(): 0, 56.1% identity in 412 aa overlap. Contains PS00096
                     Serine hydroxymethyltransferase pyridoxal-phosphate
                     attachment site and Pfam match to entry PF00464 SHMT,
```

```
                Serine hydroxymethyltransferase, score 773.50, E-value
                6.4e-232"
                /codon_start=1
                /transl_table=11
                /product="serine hydroxymethyltransferase"
                /protein_id="CAA20173.1"
                /db_xref="PID:e1314376"
                /db_xref="PID:g3402237"
                /db_xref="GI:3402237"
                /db_xref="SPTREMBL:O86565"
                /translation="MSLLNTPLHELDPDVAAAVDAELDRQQSTLEMIASENFAPVAVM
                EAQGSVLTNKYAEGYPGRRYYGGCEHVDVVEQIAIDRVKALFGAEHANVQPHSGAQAN
                AAAMFALLKPGDTIMGLNLAHGGHLTHGMKINFSGKLYNVVPYHVGDDGQVDMAEVER
                LAKETKPKLIVAGWSAYPRQLDFAAFRKVADEVGAYLMVDMAHFAGLVAAGLHPNPVP
                HAHVVTTTTHKTLGGPRGGVILSTAELAKKINSAVFPGQQGGPLEHVVAAKAVAFKVA
                ASEDFKERQGRTLEGARILAERLVRDDAKAAGVSVLTGGTDVHLVLVDLRDSELDGQQ
                AEDRLHEVGITVNRNAVPNDPRPPMVTSGLRIGTPALATRGFTAEDFAEVADVIAEAL
                KPSYDAEALKARVKTLADKHPLYPGLNK"
misc_feature    complement(3729. .4871)
                /gene="glyA"
                /note="Pfam match to entry PF00464 SHMT, Serine
                hydroxymethyltransferase, score 773.50, E-value 6.4e-232"
misc_feature    complement(4182. .4232)
                /gene="glyA"
                /note="PS00096 Serine hydroxymethyltransferase
                pyridoxal-phosphate attachment site"
gene            complement(4907. .5284)
                /gene="gcvH"
CDS             complement(4907. .5284)
                /gene="gcvH"
                /note="SC2A11.05c, gcvH, probable glycine cleavage system
                H protein, len: 125 aa; highly similar to e.g. GCSH_ECOLI
                glycine cleavage system H protein (129 aa), fasta scores;
                opt: 460 z-score: 754.1 E(): 0, 56.0% identity in 125 aa
                overlap. Contains PS00189 2-oxo acid dehydrogenases
                acyltransferase component lipoyl binding site"
                /codon_start=1
                /transl_table=11
                /product="glycine cleavage system H protein"
                /protein_id="CAA20174.1"
                /db_xref="PID:e1314377"
                /db_xref="PID:g3402238"
                /db_xref="GI:3402238"
                /db_xref="SPTREMBL:O86566"
                /translation="MSNPQQLRYSKEHEWLSGAEDGVSTVGITEHAANALGDVVFVQL
                PEVGDSVTAGETCGELESTKSVSDLYSPVSGEITEVNEDVVNDPSLVNSAPFEGGWLF
                KVRITDEPADLLSADEYTAFAGA"
misc_feature    complement(5054. .5143)
                /gene="gcvH"
                /note="PS00189 2-oxo acid dehydrogenases acyltransferase
                component lipoyl binding site"
RBS             complement(5295. .5299)
                /note="possible RBS upstream of gcvH"
gene            complement(5411. .6529)
                /gene="gcvT"
CDS             complement(5411. .6529)
                /gene="gcvT"
                /note="SC2A11.06c, gcvT, probable aminomethyltransferase
                (glycine cleavage system T protein), len: 372 aa; similar
                to e.g. GCST_BACSU probable aminomethyltransferase (EC
                2.1.2.10) (glycine cleavage system T protein) (362 aa),
                fasta scores; opt: 1074 z-score: 1108.6 E(): 0, 45.4%
                identity in 366 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="aminomethyltransferase"
                /protein_id="CAA20175.1"
                /db_xref="PID:e1314378"
                /db_xref="PID:g3402239"
                /db_xref="GI:3402239"
                /db_xref="SPTREMBL:O86567"
                /translation="MSSTELRRTALDATHRALGATMTDFAGWDMPLRYGSEREEHVAV
                RTRAGLFDLSHMGEITVTGPQAAELLNFALVGNIGTVKPGRARYTMICREDGGILDDL
                IVYRLEEAEYMVVANASNAQVVLDALTERAAGFDAEVRDDRDAYALLAVQGPESPGIL
                ASLTDADLDGLKYYAGLPGTVAGVPALIARTGYTGEDGFELFVKPEHAVGLWQALTGA
                GEAAGLIPCGLSCRDTLRLEAGMPLYGNELSTALTPFDAGLGRVVKFEKEGDFVGRAA
                LTEAAERAASRPPRVLVGLVAEGRRVPRSGYRVVAGGEVIGEVTSGAPSPTLGRPIAM
                AYVDPAHAAPGTEGVGVDIRGSHEPYEVVALPFYKRQK"
RBS             complement(6541. .6544)
                /note="possible RBS upstream of gcvT"
RBS             6942. .6946
                /note="possible RBS upstream of SC2A11.07"
gene            6956. .7624
                /gene="SC2A11.07"
CDS             6956. .7624
                /gene="SC2A11.07"
                /note="SC2A11.07, probable ATP/GTP binding protein, len:
                222 aa; contains PS00017 ATP; GTP-binding site motif A
                (P-loop)"
                /codon_start=1
                /transl_table=11
                /product="putative ATP/GTP binding protein"
                /protein_id="CAA20176.1"
                /db_xref="PID:e1314379"
                /db_xref="PID:g3402240"
                /db_xref="GI:3402240"
                /db_xref="SPTREMBL:O86568"
                /translation="MNRTTAYATTSGLALPKQPAAPAPEACAPLTAPVVRDLRDRDGR
                SPHALLFGPRDLVVITGLPGSGKSTLMRRTVRGGRIDSQDTRDRWDRRVPRLLPYTLY
                RPFVRLAHYAGLRRALRSGEGVVVHDCGTQAWVRGWLAREARRRHGTLHLLLLDVAPG
                EALAGQRERGRGVSRYAFLRHRAANSRLLRAVTEGTLPAGCSSAVVLDRAAADTLRRI
                AFTG"
misc_feature    7136. .7159
                /gene="SC2A11.07"
                /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
gene            7704. .8471
                /gene="SC2A11.08"
CDS             7704. .8471
                /gene="SC2A11.08"
                /note="SC2A11.08, unknown, len: 255 aa"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein SC2A11.08"
                /protein_id="CAA20177.1"
                /db_xref="PID:e1314380"
                /db_xref="PID:g3402241"
                /db_xref="GI:3402241"
                /db_xref="SPTREMBL:O86569"
                /translation="MDFPAQTHPHPHGGWPGNELEEVLSASLGIPSAGGRIVEVLGRS
                FLWVPLPNGGGPHSGPLDLPSLEIDGQAYVPVFSSEEQLRQVAGAHMAYTIAPAAEFA
                RGLPPQVGIAVNPEGVVGIPLPPPAVAELCRAGRTPLDGPSSGGRVRLFEPDWQDDPV
                DFLAAASAEFAATGVVRSARRCLAAVETADPVLFVGVELTHWEGDARTLPLDALGRAL
                AKSPLKYQVNLVLLEAAQDPVCDWMRERVRPFYDRDL"
gene            8666. .9463
                /gene="SC2A11.09"
CDS             8666. .9463
                /gene="SC2A11.09"
                /note="SC2A11.09, unknown, len: 265 aa"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein SC2A11.09"
                /protein_id="CAA20178.1"
                /db_xref="PID:e1314381"
                /db_xref="PID:g3402242"
                /db_xref="GI:3402242"
                /db_xref="SPTREMBL:O86570"
                /translation="MSASGTTTGQVEHMLRQVTPGRYDAYEALLRALATPTSGQVWML
                LWHGQAGSPDAQYGNMEVEGYGYAPCVTSAQELSASGWNRSYEVVDGLEAARALYPDH
                YGLWLNPHAPGGGVGIPWLDLRRIASGLDRQPAGPLRLSEPAIEIPQFYALLAQNAHR
                TPAVRALRRAWVQPALGAPYLAIGLDVYDTSPPAVDAVRAMMQQSVGAVPDGLPVSTV
                AMTDAHDPVALWLRANARPFYDREAHAPAPAQAPAGGYGYPPAGSRY"
gene            9778. .10785
                /gene="SC2A11.10"
CDS             9778. .10785
                /gene="SC2A11.10"
                /note="SC2A11.10, probable oligopeptide transport integral
```

membrane protein, len: 335 aa; similar to mnay e.g. APPC_BACSU oligopeptide transport permease protein appC (303 aa), fasta scores; opt: 465 z-score: 1038.5 E(): 0, 41.4% identity in 309 aa overlap. Also similar to S. coelicolor TR:P72406 (EMBL:U68036) bldKA (335 aa), fasta scores; opt: 912 z-score: 1192.3 E(): 0, 42.1% identity in 330 aa overlap. Contains PS00402 Binding-protein-dependent transport systems inner membrane comp signature and Pfam match to entry PF00528 BPD_transp, Binding-protein-dependent transport systems inner membrane component, score 47.00, E-value 4.1e-10"
/codon_start=1
/transl_table=11
/product="oligopeptide transport integral membrane protein"
/protein_id="CAA20179.1"
/db_xref="PID:e1314382"
/db_xref="PID:g3402243"
/db_xref="GI:3402243"
/db_xref="SPTREMBL:O86571"
/translation="MTAPLHEPTAEAAPSAAEEAAVAGAEAKSVQGRSLGRIAWERLK
RDKLALAGGIVVLVLIVVAVFAPLITSLYGQDPNAYNEDMIDPLFGTPTGSLGGLGGE
HWLGVEPVNGRDIFARIVHGARISLLVGFLSAMVAVVLGTVLGVLAGFFGGWVDSLIC
RIMDGLLAFPQLLFIIALVSVMPDNMLGLSGTGVRLLMMIVVIGFFGWPYIGRVVRGQ
TLSMREREYVEAARSLGAGRFYILFKELLPNLVAPIIVYTTMMIPTNILTEAALSFLG
VGVKPPTSSWGQMLSSAIDYYKSDPMYMVVPGVAIFITVLAFNLFGDGVRDALDPKGS
R"

misc_feature 10432. .10662
/gene="SC2A11.10"
/note="Pfam match to entry PF00528 BPD_transp, Binding-protein-dependent transport systems inner membrane component, score 47.00, E-value 4.1e-10"

misc_feature 10435. .10521
/gene="SC2A11.10"
/note="PS00402 Binding-protein-dependent transport systems inner membrane comp signature"

RBS 10842. .10846
/note="possible RBS upstream of SC2A11.11"

gene 10857. .12659
/gene="SC2A11.11"

CDS 10857. .12659
/gene="SC2A11.11"
/note="SC2A11.11, probable oligopeptide-binding lipoprotein, len: 600 aa; similar to e.g. OPPA_BACSU oligopeptide-binding protein oppA precursor (545 aa), fasta scores; opt: 401 z-score: 242.5 E(): 3e-06, 25.5% identity in 554 aa overlap Also similar to S. coelicolor TR:P72407 (EMBL:U68036) bldKB (602 aa), fasta scores; opt: 702 z-score: 583.5 E(): 3e-25, 29.3% identity in 614 aa overlap. Contains N-terminal signal sequence and appropriately positioned PS00013 Prokaryotic membrane lipoprotein lipid attachment site, and Pfam match to entry PF00496 SBP_bac_5, Bacterial extracellular solute-binding proteins, family 5, score 45.60, E-value 7.2e-11"
/codon_start=1
/transl_table=11
/product="putative oligopeptide-binding lipoprotein"
/protein_id="CAA20180.1"
/db_xref="PID:e1314383"
/db_xref="PID:g3402244"
/db_xref="GI:3402244"
/db_xref="SPTREMBL:O86572"
/translation="MTTQRTSGRRKQALAAAAVVAALLTTAACGGGGDDDSGGGSKNG
AAGYDAANNKVAQASEAKKGGTLKFASTQDADSWDTTRGYYGFMWDFSRYYSRQLVTN
KTEPGAAGAETTPDLATDVAKVSEDGKTYTYTLRDGVTWEDGKPITSKDVKYGIERVW
AQDVLSGGPTYLKEVLDPKGEYQGPYKDTSKDKLGLKAIETPDDKTIVFHLPQANSDF
EEMLALTSASPVRQDKDTKSKYGLHPFSSGPYKFESYNPGKDLTLVRNTEWKQASDPV
RKAYPDKITIKFFTNANDMDARLIAGDYDIDLAQTGLSPQGRTTALKEHKANLDNPVS
GYIRYAAFPQNVKPFDNIHCRKAVLYGADHVSLQTARGGPVAGGDIGTNMLPPAVPGS
EGQKYDPYEMSGANKKGNVEKAKEELKACNQPNGFKTTIAVRNNKPVEVATAESLQAS
LKKVGIDVEIDQYDGSQYASVIGSPSNVQKKGYGIIIMGWGPDFPSVQGYGLPLWHSD
YILDSANNNFAMIDDPKIDGLFADYLTAKDDAGKKQIATEINHKVMEGAYYLPFVFEK
FINWRSSNIANVYTTDAYSGMYDFVNIGLKNPKK"

misc_feature 10911. .10943
/gene="SC2A11.11"
/note="PS00013 Prokaryotic membrane lipoprotein lipid attachment site"

misc_feature 11181. .11309
/gene="SC2A11.11"
/note="Pfam match to entry PF00496 SBP_bac_5, Bacterial extracellular solute-binding proteins, family 5, score 45.60, E-value 7.2e-11"

gene 12768. .13766
/gene="SC2A11.12"

CDS 12768. .13766
/gene="SC2A11.12"
/note="SC2A11.12, probable oligopeptide transport system integral membrane protein, len: 332 aa; similar to many e.g. OPPB_ECOLI oligopeptide transport system permease protein (306 aa), fasta scores; opt: 568 z-score: 807.4 E(): 0, 35.0% identity in 337 aa overlap Also similar to S. coelicolor TR:P72408 (EMBL:U68036) bldKC (323 aa), fasta scores; opt: 875 z-score: 1186.1 E(): 0 42.4% identity in 337 aa overlap. Contains PS00402 Binding-protein-dependent transport systems inner membrane comp signature and Pfam match to entry PF00528 BPD_transp, Binding-protein-dependent transport systems inner membrane component, score 34.10, E-value 3.1e-06"
/codon_start=1
/transl_table=11
/product="oligopeptide transport system integral membrane protein"
/protein_id="CAA20181.1"
/db_xref="PID:e1314384"
/db_xref="PID:g3402245"
/db_xref="GI:3402245"
/db_xref="SPTREMBL:O86573"
/translation="MLAYLIRRLFAAAVMLVVIILVVFCIFFLVPKWAGVDIALNFVG
KQADPAAVEGVREKLGLGDPVFVQAWEFFKGIFAGRTYAAGGDVTHCAAPCFGYSFKT
EQSVWPVLTERFPVTLALALGAAVLWLIFGVAAGVLSALKRGTLWDRGAMVVALAGVS
LPIYFTGLLSLAIFSYGLGWIDGEFVPLEDSLTGWLGGMILPWITLAFLYAAMYARIT
RATMLEILGEDYIRTARAKGLKEQVVISKHAMRSTLTPLLTMLGMDLGALMGGAILTE
TTFSLPGLGQKVLDAIKNHDLPFILGVVLITSLAVLIANLVVDILYAVIDPRVRLA"

misc_feature 13431. .13646
/gene="SC2A11.12"
/note="Pfam match to entry PF00528 BPD_transp, Binding-protein-dependent transport systems inner membrane component, score 34.10, E-value 3.1e-06"

misc_feature 13434. .13520
/gene="SC2A11.12"
/note="PS00402 Binding-protein-dependent transport systems inner membrane comp signature"

RBS 13754. .13757
/gene="SC2A11.12"
/note="possible RBS upstream of SC2A11.13"

gene 13763. .14860
/gene="SC2A11.13"

CDS 13763. .14860
/gene="SC2A11.13"
/note="SC2A11.13, probable oligopeptide ABC transporter ATP-binding protein, len: 365 aa; similar to many e.g. OPPD_BACSU oligopeptide transport ATP-binding protein (336 aa), fasta scores; opt: 1089 z-score: 1413.1 E(): 0, 52.3% identity in 323 aa overlap Also similar to S. coelicolor TR:P72409 (EMBL:U68036) bldKD (353 aa), fasta scores; opt: 1396 z-score: 1731.9 E(): 0, 58.4% identity in 365 aa overlap. Contains PS00017 ATP /GTP-binding site motif A (P-loop), PS00211 ABC transporters family signature and Pfam match to entry PF00005 ABC_tran, ABC transporters, score 203.50, E-value 3.3e-57"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter ATP-binding protein"
/protein_id="CAA20182.1"
/db_xref="PID:e1314385"

```
                /db_xref="PID:g3402246"
                /db_xref="GI:3402246"
                /db_xref="SPTREMBL:O86574"
                /translation="MTELSKSGAAVGEPTGSSPAPTAFLEVRDLKVHFPTDDGLVKSV
                DGLSFQLEKGKTLGIVGESGSGKSVTSLGIMGLHTAGQYGKRKARISGEIWLDGTELL
                SADPDHVRKLRGREMAMIFQDPLSALHPYYTIGQQIVEAYRIHHKVDKKTAKRRAVEM
                LDRVGIPQPDKRVDSYPHEFSGGMRQRAMIAMSLVNNPELLIADEPTTALDVTVQAQI
                LDLIRDLQKEFGSAVIVITHDLGVVAELADDLLVMYGGRCVERGPAEKVFYEPRHPYT
                WGLLGSMPRLDRDQQERLIPVKGSPPSLINLPSGCAFNPRCPYADVPKDDVTRTVRPE
                LTEVGSKHWAACHMSQEQRERIWTEEIAPKL"
misc_feature    13922. .14530
                /gene="SC2A11.13"
                /note="Pfam match to entry PF00005 ABC_tran, ABC
                transporters, score 203.50, E-value 3.3e-57"
misc_feature    13943. .13966
                /gene="SC2A11.13"
                /note="PS00017 ATP /GTP-binding site motif A (P-loop)"
misc_feature    14300. .14344
                /gene="SC2A11.13"
                /note="PS00211 ABC transporters family signature"
gene            14857. .16347
                /gene="SC2A11.14"
CDS             14857. .16347
                /gene="SC2A11.14"
                /note="SC2A11.14, probable oligopeptide ABC transporter
                ATP-binding protein, len: 496 aa; similar to many e.g.
                APPF_BACSU oligopeptide transport ATP-binding protein (329
                aa), fasta scores; opt: 1284 z-score: 1056.3 E(): 0, 56.4%
                identity in 326 aa overlap Also similar to S. coelicolor
                TR:P72410 (EMBL:U68036) bldKE (284 aa), fasta scores; opt:
                1114 z-score: 1091.6 E(): 0, 59.3% identity in 280 aa
                overlap. Contains PS00017 ATP /GTP-binding site motif A
                (P-loop), PS00211 ABC transporters family signature and
                Pfam match to entry PF00005 ABC_tran, ABC transporters,
                score 202.90, E-value 4.8e-57. C-terminus contains 13
                degenerate repeats of AS(D /S)DA"
                /codon_start=1
                /transl_table=11
                /product="oligopeptide ABC transporter ATP-binding
                protein"
                /protein_id="CAA20183.1"
                /db_xref="PID:e1314386"
                /db_xref="PID:g3402247"
                /db_xref="GI:3402247"
                /db_xref="SPTREMBL:O86575"
                /translation="MSDEAKDEAVAIPAQSDGTESTSAGAATLTKGAAPGETLLKVTG
                LQKHFPIRKGLLQRQVGAVRAVDGIDFEVRSGETLGVVGESGCGKSTMGRLITRLLEP
                TAGKVEFQGKDITHLGVGGMRPLRRDVQMIFQDPYSSLNPRHTIGTIVSAPFKLQGVQ
                PEGGVKKEVQRLLSVVGLNPEHYNRYPHEFSGGQRQRIGIARALALNPRLVVADEPVS
                ALDVSIQAQVVNLLDDLQQELGLTYVIIAHDLSVVRHVSDRIAVMYLGKIMELADRDL
                LYKTPMHPYTKALMSAVPIPDPARRGAKSERILLKGDVPSPIAPPSGCRFHTRCWKAT
                QVCTTTEPPLKELRPGQQVACHHPENFEDQAPQDVVLLTAAKEAAELVSEDALAESAA
                TSAAVAAEVEASEAEASEETASSDAASDGAASEEAASSDAASDDSASDGAAPKEAASS
                DAASDDAGSSDSVSDDAAPDDSASDGDEPAAESAGSTDEEPPAKEK"
misc_feature    15082. .15654
                /gene="SC2A11.14"
                /note="Pfam match to entry PF00005 ABC_tran, ABC
                transporters, score 202.90, E-value 4.8e-57"
misc_feature    15103. .15126
                /gene="SC2A11.14"
                /note="PS00017 ATP /GTP-binding site motif A (P-loop)"
misc_feature    15424. .15468
                /gene="SC2A11.14"
                /note="PS00211 ABC transporters family signature"
repeat_region   16093. .16287
                /note="13x15bp degenerate repeats"
gene            16425. .17084
                /gene="SC2A11.15"
CDS             16425. .17084
                /gene="SC2A11.15"
                /note="SC2A11.15, probable integral membrane protein, len:
                219 aa; similar to hypothetical proteins e.g. YADS_ECOLI
                (207 aa), fasta scores; opt: 397 z-score: 543.4 E():
                5.2e-23, 36.7% identity in 196 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="putative integral membrane protein"
                /protein_id="CAA20184.1"
                /db_xref="PID:e1314387"
                /db_xref="PID:g3402248"
                /db_xref="GI:3402248"
                /db_xref="SPTREMBL:O86576"
                /translation="MYEQLFSPTVQHTLDLVGIFVFAISGALLAVRKNFDVFGIAVLA
                EVTALGGGLFRDLVIGAVPPAAFTDLGYFLTPLLATLLVFFLHPHVERLQTGVNIFDA
                AGLGLFCVAGTTKAYDYGLGLTASACLGLTTAVGGGVLRDVLANEVPSLLRWDRDLYA
                VPAIVGSAMVALCIRYEALTPFTSGLAVVTAFVLRLLALRFHWRAPRAWNRRSTVVEG
                D"
gene            17220. .18242
                /gene="SC2A11.16"
CDS             17220. .18242
                /gene="SC2A11.16"
                /note="SC2A11.16, unknown, len: 340 aa; some similarity to
                M. tuberculosis hypothetical protein YV25_MYCTU
                MTCY20G9.25, Rv0499 (291 aa), fasta scores; opt: 240
                z-score: 316.7 E(): 2.2e-10, 28.9% identity in 270 aa
                overlap"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein SC2A11.16"
                /protein_id="CAA20185.1"
                /db_xref="PID:e1314388"
                /db_xref="PID:g3402249"
                /db_xref="GI:3402249"
                /db_xref="SPTREMBL:O86577"
                /translation="MEPLEAALREWHDRVSARGRAMATGSARLESCPRAREERERKLP
                LSSFLLYRAVMPEAASAAAVRVTIGDSEFDRDTAVTRRAPGVYDIDLSAGWTIISAVN
                GGYLLAVLGRALGDALPHPDPFTISAHYLTASRPGPAVVRTETVRTGRSLSTGQASLL
                QYDDEGNEVERIRVLASYGDLDALPDDVRTSAKPPAMPPMDQCFGPEDGPAPVEGSSA
                IADRLMLKLDPATLGWALGQPSGRGEMRAWFGLADGRDADPLSLLLAVDALPPTAFEL
                GLKGWVPTVELTAHVRCRPAPGPLRVSITTRNLAGGFLEEDAEVWDSTDRLVAQSRQL
                ARVRLG"
gene            complement(18271. .18897)
                /gene="SC2A11.17c"
CDS             complement(18271. .18897)
                /gene="SC2A11.17c"
                /note="SC2A11.17c, probable transcriptional regulator,
                len: 208 aa; similar to members of the tetR family e.g.
                TCMR_STRGA tetracenomycin C transcriptional repressor (226
                aa), fasta scores; opt: 138 z-score: 224.5 E(): 3e-05,
                26.5% identity in 155 aa overlap. Contains
                helix-turn-helix motif at aa 36-57 (Score 1332, +3.72 SD)
                and Pfam match to entry PF00440 tetR, Bacterial regulatory
                proteins, tetR family, score 23.40, E-value 9.9e-05"
                /codon_start=1
                /transl_table=11
                /product="putative transcriptional regulator"
                /protein_id="CAA20186.1"
                /db_xref="PID:e1314389"
                /db_xref="PID:g3402250"
                /db_xref="GI:3402250"
                /db_xref="SPTREMBL:O86578"
                /translation="MSHTPGVRQAQKRKTRQALLDAALGLLEEQSLSSLGLREVTRAV
                GVAPTAFYRHFHSTADLGVALVEEALGSLHPMIRTTMSPAEDDDERIARAVELIAGHV
                AAHPAHVRFIARERHGGVQPVRRAIREQLARFAEEVRAELATDASAAGWSETDLRMLA
                GLYVDQMLITASLFLETLDAPEEERRRAVETARRQLRLISVGRTHWLD"
misc_feature    complement(18733. .18843)
                /gene="SC2A11.17c"
                /note="Pfam match to entry PF00440 tetR, Bacterial
                regulatory proteins, tetR family, score 23.40, E-value
                9.9e-05"
                /product="putative transcriptional regulator"
RBS             19011. .19015
                /note="RBS upstream of SC2A11.18"
gene            19025. .19192
                /gene="SC2A11.18"
CDS             19025. .19192
```

```
                     /gene="SC2A11.18"
                     /note="SC2A11.18, small hydrophobic protein, len: 55 aa"
                     /codon_start=1
                     /transl_table=11
                     /product="small hydrophobic protein"
                     /protein_id="CAA20187.1"
                     /db_xref="PID:e1314390"
                     /db_xref="PID:g3402251"
                     /db_xref="GI:3402251"
                     /db_xref="SPTREMBL:O86579"
                     /translation="MLFLVVALLLLGVVLGAVAHAPLTFSVAAGLVIAVWLGIFAVRE
                     RHGRRRGTSAH"
     RBS             19217. .19220
                     /note="possible RBS upstream of SC2A1.19"
     gene            19227. .19520
                     /gene="SC2A11.19"
     CDS             19227. .19520
                     /gene="SC2A11.19"
                     /note="SC2A11.19, small hydrophobic protein, len: 97 aa"
                     /codon_start=1
                     /transl_table=11
                     /product="small hydrophobic protein"
                     /protein_id="CAA20188.1"
                     /db_xref="PID:e1314391"
                     /db_xref="PID:g3402252"
                     /db_xref="GI:3402252"
                     /db_xref="SPTREMBL:O86580"
                     /translation="MQLTAPADRTTESPRTTRRTGIHDTDGMAVASFVLGLLGLLVLN
                     VFLGPVAIALAGIALWRGTARRGRALLGLGLGVADLLVLVAFMSADSTLSWSL"
     gene            19611. .20780
                     /gene="SC2A11.20"
     CDS             19611. .20780
                     /gene="SC2A11.20"
                     /note="SC2A11.20, probable pyridoxal-phosphate-dependent
                     aminotransferase, len: 389 aa; similar to many e.g.
                     NIFS_ANASP Anaebaena nifS protein (400 aa), fasta scores;
                     opt: 855 z-score: 1070.8 E(): 0, 40.4% identity in 384 aa
                     overlapand SPL1_CANAL Candida albicans Trna splicing
                     protein SPL1 (488 aa), fasta scores; opt: 791 z-score:
                     912.8 E(): 0, 36.6% identity in 388 aa overlap. Contains
                     Pfam match to entry PF00266 aminotran_5, Aminotransferases
                     class-V, score 259.00, E-value 6.2e-74"
                     /codon_start=1
                     /transl_table=11
                     /product="putative pyridoxal-phosphate-dependent
                     aminotransferase"
                     /protein_id="CAA20189.1"
                     /db_xref="PID:e1314392"
                     /db_xref="PID:g3402253"
                     /db_xref="GI:3402253"
                     /db_xref="SPTREMBL:O86581"
                     /translation="MAYLDHAATTPMLPEAVEALTAHLSATGNASSLHASGRRARRTV
                     EEARETLAEALGARPSEVVLTSGGTEADNLAVKGLYWARRDADPARTRVLASPVEHHA
                     VLDAVHWLGEHEGATVEYLPVDAQGRVHPEALREAIARDPGNVALATVMWANNEIGTI
                     MPVRELADVAAEFDVPLHADAVQAFGQLPVGFADSGLAAMTVSGHKIGGPYGIGALLL
                     GRDYTPVPVLHGGGQERHVRSGTLDVPAVASFAVAGRLAAEQREWFAREIGALRDDLV
                     DAVRTAVPDAILGGDPSPGGRLPANAHFTFPGCEGDSLLLLDAQGIECSTGSACTAG
                     VAQPSHVLLATGTDPDLARGTLRFSLGHTSTEADVEAVAKAIGPAVERARAAGLS"
     misc_feature    19644. .20726
                     /gene="SC2A11.20"
                     /note="Pfam match to entry PF00266 aminotran_5,
                     Aminotransferases class-V, score 259.00, E-value 6.2e-74"
     gene            complement(20817. .21545)
                     /gene="SC2A11.21c"
     CDS             complement(20817. .21545)
                     /gene="SC2A11.21c"
                     /note="SC2A11.21c, unknown, len: 242 aa; some simialrity
                     in C-terminus to enterobacterial ampD genes e.g.
                     AMPD_ECOLI AMPD protein (183 aa), fasta scores; opt: 210
                     z-score: 212.6 E(): 0.00014, 31.5% identity in 149 aa
                     overlapand Enterobacter cloacae TR:Q46957 (EMBL:U40785)
                     anhydro-n-acetylmuramyl-tripeptide amidase (187 aa), fasta
                     scores; opt: 207 z-score: 192.0 E(): 0.0019, 31.6%
                     identity in 155 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein SC2A11.21c"
                     /protein_id="CAA20190.1"
                     /db_xref="PID:e1314393"
                     /db_xref="PID:g3402254"
                     /db_xref="GI:3402254"
                     /db_xref="SPTREMBL:O86582"
                     /translation="MGAKQGSTSGSTPGGPAGGGSADGDRRIGRRALIVGGAAAAVGT
                     GVLARDELSRLWWRIPGVERPREDGVVDYAGARWVAASDANWRRADRPDDFGIDMVIV
                     HVTQGSFDSAVRAFQDPGHKAAAHYIVGQDGRVTQMIRELDVAYHAGNRDYNERSVGI
                     EHAGFVDRPQDFTDAMYAASARLTAGICARYGIPVDRKHILGHVEVPGTDHTDPGPHW
                     DWDRYLELVRRAATSGPSKPPTKA"
     RBS             21598. .21602
                     /note="possible RBS upstream of trmU"
     gene            21607. .22737
                     /gene="trmU"
     CDS             21607. .22737
                     /gene="trmU"
                     /note="SC2A11.22, trmU, probable tRNA
                     (5-methylaminomethyl-2-thiouridylate)- methyltransferase,
                     len: 376 aa; similar to many e.g. TRMU_BACSU probable tRNA
                     (5-methylaminomethyl-2-thiouridylate)- methyltransferase
                     (370 aa), fasta scores; opt: 708 z-score: 748.3 E(): 0,
                     38.1% identity in 365 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="tRNA (5-methylaminomethyl-2-
                     thiouridylate)-methyltransferase"
                     /protein_id="CAA20191.1"
                     /db_xref="PID:e1314394"
                     /db_xref="PID:g3402255"
                     /db_xref="GI:3402255"
                     /db_xref="SPTREMBL:O86583"
                     /translation="MTETPQRTRPLRVLAAMSGGVDSAVAAARAAEAGHDVTGVHLAL
                     SANPQSFRTGARGCCTIEDSRDARRAADVIGIPFYVWDLADRFREDVVEDFVAEYEAG
                     RTPNPCLRCNEKIKFAALLDKALALGFDAVCTGHYAQVILREDGVRELHRASDMAKDQ
                     SYVLGVLDDRQLAHAMFPLGDTVTTKDEIRAEAERRGLAVAKKPDSHDICFIADGNTQ
                     GFLADRLGKAEGDIVDEAGNRLGTHEGAYGYTIGQRKGLRIGTPAPDGKPRYVLDISP
                     VNNTVTVGPAEALDVDALRAIKPRWCGAAPTGPGTYTAQLRAHGGETEVRAELVDGTL
                     EVTFSEPVRGVAPGQAIVLYDGTRVVGSATIASTTRATAGAA"
     misc_feature    complement(22656. .22789)
                     /note="nominal overlap with cosmid 8D9 from 17389 to
                     17522"
BASE COUNT     3379 a   8444 c   7708 g   3258 t
ORIGIN


  Query Match           3.8%;  Score 37.6;  DB 1;  Length 22789;
  Best Local Similarity   55.3%;  Pred. No. 12;
  Matches   73;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy    110 ACCTGTGGCTCCTGGCCTGCCTGGTGGCCGGCTTCCTGGGAGCCTGGGCCCCCGCTGTCC 169
          | |||      ||  | |  | || | || | |  | | | |||   ||||   |  | ||
Db   4663 ATCTGCTCGACCACGTCGACGTGCTCGCAGCCGCCGTAGTAGCGGCGGCCGGGGTAGCCC 4722

Qy    170 ACGCCCAAGGTGTCTTTGAGGACTGCTGCCTGGCCTACCACTACCCCATTGGGTGGGCTG 229
           || |  |  |||   |||||||  |   ||||| |  |||  ||| |    || | |  |
Db   4723 TCGGCGTACTTGTTGGTGAGGACCGAGCCCTGGGCCTCCATGACCGCGACCGGGGCGAAG 4782

Qy    230 TGCTCCGGCGCG 241
          | ||||| ||||
Db   4783 TTCTCCGACGCG 4794


RESULT  12
AC007387/c
LOCUS       AC007387   199248 bp    DNA             HTG       05-JUN-1999
DEFINITION  Homo sapiens clone NH0372J12, WORKING DRAFT SEQUENCE, 2 unordered
            pieces.
ACCESSION   AC007387
NID         g5001502
```

```
VERSION     AC007387.2  GI:5001502
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 199248)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 199248)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 5, 1999 this sequence version replaced gi:4680449.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    63149: contig of 63149 bp in length
            *   63150    63167: gap of unknown length
            *   63168   199248: contig of 136081 bp in length.
FEATURES             Location/Qualifiers
     source          1. .199248
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NH0372J12"
BASE COUNT    58464 a  43548 c  41243 g  55975 t     18 others
ORIGIN
```

Query Match 3.8%; Score 37.6; DB 35; Length 199248;
Best Local Similarity 50.6%; Pred. No. 12;
Matches 91; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

```
Qy      70 CGTCCAGGTGCCCAGGGAGGAGGACCCGCCTGCAGCATGAACCTGTGGCTCCTGGCCTGC 129
           | |||    |  | ||  ||  | | | | | |  ||| ||     ||||   || |
Db  79901 CCTCCCTCCGAACGGGACGGCTGGCTGGGCGGGGGGCTGACCCCCCACCTCCCTCCCGGA 79842

Qy     130 CTGGTGGCCGGCTTCCTGGGAGCCTGGGCCCCCGCTGTCCACGCCCAAGGTGTCTTTGAG 189
             ||  | | |  |  |||| | |||  |||||    |||   ||  | | | |     |
Db  79841 TGGGGCGGCTGGCTGGTGGGGGGCTGACCCCCCCACCTCCTTCCCGGATGAGGCGGCTGG 79782

Qy     190 GACTGCTGCCTGGCCTACCACTACCCCATTGGGTGGGCTGTGCTCCGGCGCGCCTGGACT 249
             || |   |||    |||| |||  | |||  ||||| ||   ||||| |||   ||
Db  79781 CTGGGCAGAGGGGCTCCTCACTTCCCAGTAGGGGCGGCTGGGCAGAGGCGCCCCTCACCT 79722
```

```
RESULT  13
CFY15483
LOCUS       CFY15483     4702 bp    mRNA            MAM       17-JUL-1998
DEFINITION  Canis familiaris mRNA for retinal guanylate cyclase E.
ACCESSION   Y15483
NID         g2648063
VERSION     Y15483.1  GI:2648063
KEYWORDS    guanylate cyclase E.
SOURCE      dog.
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 4702)
  AUTHORS   Veske,A., Nilsson,S.E. and Gal,A.
  TITLE     Organization of the canine gene encoding the E isoform of retinal
            guanylate cyclase (cGC-E) and exclusion of its involvement in the
            inherited retinal dystrophy of the Swedish Briard and Briard-beagle
            dogs
  JOURNAL   Biochim. Biophys. Acta 1372 (1), 69-77 (1998)
  MEDLINE   98316153
REFERENCE   2  (bases 1 to 4702)
  AUTHORS   Veske,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-1997) A. Veske, Universitaets-Krankenhaus
            Eppendorf, Institut fuer Humangenetik, Martinistrasse 52, D- 20246
            Hamburg, FRG
FEATURES             Location/Qualifiers
     source          1. .4702
                     /organism="Canis familiaris"
                     /strain="Beagle / Briard cross"
                     /isolate="Drutten"
                     /db_xref="taxon:9615"
                     /dev_stage="adult"
                     /tissue_type="retina"
     CDS             183. .3512
                     /codon_start=1
                     /product="guanylate cyclase E"
                     /protein_id="CAA75655.1"
                     /db_xref="PID:e1187370"
                     /db_xref="PID:g2648064"
                     /db_xref="GI:2648064"
                     /db_xref="SWISS-PROT:O19179"
                     /translation="MSACALLAGGLPDPRLCAPARWARSPPGVPGAPPWPQPRLRLLL
                     LLLLLPPSALSAVFTVGVLGPWACDPIFARARPDLAARLAAARLNRDAALEDGPRFEV
                     TLLPEPCRTPGSLGAVSSALGRVSGLVGPVNPAACRPAELLAQEAGVALVPWSCPGTR
                     AGGTTAPAGTPAADALYALLRAFRWARVALITAPQDLWVEAGRALSAALRARGLPVAL
                     VTTMEPSDLSGAREALRRVQDGPRVRAVIMVMHSVLLGGEEQRCLLQAAEELGLADGS
                     LVFLPFDTLHYALSPGPEALAVLANSSQLRRAHDAVLILTRHCPPGGSVMDNLRRAQE
                     HQELPSDLDLQQVSPFFGTIYDAVLLLAGGVARARAAAGGGWVSGATVAHHIPDAQVP
                     GFCGTLGGAQEPPFVLLDTDAAGDRLFATYMLDPTRGSLLSAGTPVHFPRGGGTPGSD
                     PSCWFEPGVICNGGVEPGLVFLGFLLVVGMGLTGAFLAHYLRHRLLHIQMVSGPNKII
                     LTLDDVTFLHPHGGSTRKVVQGSRSSLAARSTSDIRSVPSQPLDNSNIGLFEGDWVWL
                     KKFPGDQHIAIRPATKTAFSKLRELRHENVVLYLGLFLGSGGAGGSAAGEGVLAVVSE
                     HCARGSLHDLLAQRDIKLDWMFKSSLLLDLIKGMRYLHHRGVAHGRLKSRNCVVDGRF
                     VLKVTDHGHARLMEAQRVLLEPPSAEDQLWTAPELLRDPALERRGTLPGDVFSLGIIM
                     QEVVCRSAPYAMLELTPEEVVERVRSPPPLCRPSVSMDQAPVECIQLMKQCWAEHPDL
                     RPSLGHIFDQFKSINKGRKTNIIDSMLRMLEQYSSNLEDLIRERTEELELEKQKTDRL
                     LTQMLPPSVAEALKMGTPVEPEYFEEVTLYFSDIVGFTTISAMSEPIEVVDLLNDLYT
                     LFDAIIGSHDVYKVETIGDAYMVASGLPQRNGQRHAAEIANMALDILSAVGSFRMRHM
                     PEVPVRIRIGLHSGPCVAGVVGLTMPRYCLFGDTVNTASRMESTGLPYRIHVNMSTVR
                     ILHALDEGFQTEVRGRTELKGKGAEDTYWLVGRRGFNKPIPKPPDLQPGASNHGISLQ
                     EIPLDRRWKLEKARPGQFSGK"
BASE COUNT      821 a   1528 c   1499 g    854 t
ORIGIN
```

Query Match 3.8%; Score 37.4; DB 3; Length 4702;
Best Local Similarity 58.6%; Pred. No. 13;
Matches 65; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```
Qy      54 CGGTGGCCCCGTTATTCGTCCAGGTGCCCAGGGAGGAGGACCCGCCTGCAGCATGAACCT 113
           ||  | || | | |  ||    |  |||| || ||  || |||  || |    ||| ||
Db    434 CGCCGCCCGCCTGAACCGCGACGCAGCCCTGGAAGACGGGCCCCGCTTCGAGGTGACGCT 493

Qy     114 GTGGCTCCTGGCCTGCCTGGTGGCCGGCTTCCTGGGAGCCTGGGCCCCCGC 164
           |  || |  | | |||| |  | | |||| ||||| |||  | || ||||
Db    494 GCTGCCCGAGCCATGCCGGACGCCGGGCTCCCTGGGGGCCGTGTCCTCCGC 544
```

```
RESULT  14
CFY15484
LOCUS       CFY15484     14515 bp    DNA             MAM       17-JUL-1998
DEFINITION  Canis familiaris gene encoding retinal guanylate cyclase E.
ACCESSION   Y15484
NID         g2648065
VERSION     Y15484.1  GI:2648065
KEYWORDS    guanylate cyclase E.
SOURCE      dog.
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 14515)
```

```
  AUTHORS   Veske,A., Nilsson,S.E. and Gal,A.
  TITLE     Organization of the canine gene encoding the E isoform of retinal
            guanylate cyclase (cGC-E) and exclusion of its involvement in the
            inherited retinal dystrophy of the Swedish Briard and Briard-beagle
            dogs
  JOURNAL   Biochim. Biophys. Acta 1372 (1), 69-77 (1998)
  MEDLINE   98316153
REFERENCE   2  (bases 1 to 14515)
  AUTHORS   Veske,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-1997) A. Veske, Universitaets-Krankenhaus
            Eppendorf, Institut fuer Humangenetik, Martinistrasse 52, D- 20246
            Hamburg, FRG
FEATURES             Location/Qualifiers
     source          1. .14515
                     /organism="Canis familiaris"
                     /strain="Beagle / Briard cross"
                     /isolate="2 briards"
                     /db_xref="taxon:9615"
                     /dev_stage="adult"
                     /tissue_type="leukocytes"
     mRNA            join(1. .112,392. .1194,1289. .1593,4284. .4635,5016. .5100,
                     5335. .5437,5829. .5930,7432. .7512,8685. .8891,8979. .9135,
                     9579. .9728,10324. .10472,10957. .11120,11212. .11404,
                     11621. .11795,12044. .12142,12260. .12354,12475. .12560,
                     12689. .12798,13358. .14515)
     exon            1. .112
                     /number=1
     intron          113. .391
                     /number=1
     exon            392. .1194
                     /number=2
     CDS             join(462. .1194,1289. .1593,4284. .4635,5016. .5100,
                     5335. .5437,5829. .5930,7432. .7512,8685. .8891,8979. .9135,
                     9579. .9728,10324. .10472,10957. .11120,11212. .11404,
                     11621. .11795,12044. .12142,12260. .12354,12475. .12560,
                     12689. .12782)
                     /codon_start=1
                     /product="guanylate cyclase E"
                     /protein_id="CAA75656.1"
                     /db_xref="PID:e1187372"
                     /db_xref="PID:g2648066"
                     /db_xref="GI:2648066"
                     /db_xref="SWISS-PROT:O19179"
                     /translation="MSACALLAGGLPDPRLCAPARWARSPPGVPGAPPWPQPRLRLLL
                     LLLLLPPSALSAVFTVGVLGPWACDPIFARARPDLAARLAAARLNRDAALEDGPRFEV
                     TLLPEPCRTPGSLGAVSSALGRVSGLVGPVNPAACRPAELLAQEAGVALVPWSCPGTR
                     AGGTTAPAGTPAADALYALLRAFRWARVALITAPQDLWVEAGRALSAALRARGLPVAL
                     VTTMEPSDLSGAREALRRVQDGPRVRAVIMVMHSVLLGGEEQRCLLQAAEELGLADGS
                     LVFLPFDTLHYALSPGPEALAVLANSSQLRRAHDAVLILTRHCPPGGSVMDNLRRAQE
                     HQELPSDLDLQQVSPFFGTIYDAVLLLAGGVARARAAAGGGWVSGATVAHHIPDAQVP
                     GFCGTLGGAQEPPFVLLDTDAAGDRLFATYMLDPTRGSLLSAGTPVHFPRGGGTPGSD
                     PSCWFEPGVICNGGVEPGLVFLGFLLVVGMGLTGAFLAHYLRHRLLHIQMVSGPNKII
                     LTLDDVTFLHPHGGSTRKVVQGSRSSLAARSTSDIRSVPSQPLDNSNIGLFEGDWVWL
                     KKFPGDQHIAIRPATKTAFSKLRELRHENVVLYLGLFLGSGGAGGSAAGEGVLAVVSE
                     HCARGSLHDLLAQRDIKLDWMFKSSLLLDLIKGMRYLHHRGVAHGRLKSRNCVVDGRF
                     VLKVTDHGHARLMEAQRVLLEPPSAEDQLWTAPELLRDPALERRGTLPGDVFSLGIIM
                     QEVVCRSAPYAMLELTPEEVVERVRSPPPLCRPSVSMDQAPVECIQLMKQCWAEHPDL
                     RPSLGHIFDQFKSINKGRKTNIIDSMLRMLEQYSSNLEDLIRERTEELELEKQKTDRL
                     LTQMLPPSVAEALKMGTPVEPEYFEEVTLYFSDIVGFTTISAMSEPIEVVDLLNDLYT
                     LFDAIIGSHDVYKVETIGDAYMVASGLPQRNGQRHAAEIANMALDILSAVGSFRMRHM
                     PEVPVRIRIGLHSGPCVAGVVGLTMPRYCLFGDTVNTASRMESTGLPYRIHVNMSTVR
                     ILHALDEGFQTEVRGRTELKGKGAEDTYWLVGRRGFNKPIPKPPDLQPGASNHGISLQ
                     EIPLDRRWKLEKARPGQFSGK"
     intron          1195. .1288
                     /number=2
     exon            1289. .1593
                     /number=3
     intron          1594. .4283
                     /number=3
     exon            4284. .4635
                     /number=4
     intron          4636. .5015
                     /number=4
     exon            5016. .5100
                     /number=5
     intron          5101. .5334
                     /number=5
     exon            5335. .5437
                     /number=6
     intron          5438. .5828
                     /number=6
     exon            5829. .5930
                     /number=7
     intron          5931. .7431
                     /number=7
     exon            7432. .7512
                     /number=8
     intron          7513. .8684
                     /number=8
     exon            8685. .8891
                     /number=9
     intron          8892. .8978
                     /number=9
     exon            8979. .9135
                     /number=10
     intron          9136. .9578
                     /number=10
     exon            9579. .9728
                     /number=11
     intron          9729. .10323
                     /number=11
     exon            10324. .10472
                     /number=12
     intron          10473. .10956
                     /number=12
     exon            10957. .11120
                     /number=13
     intron          11121. .11211
                     /number=13
     exon            11212. .11404
                     /number=14
     intron          11405. .11620
                     /number=14
     exon            11621. .11795
                     /number=15
     intron          11796. .12043
                     /number=15
     exon            12044. .12142
                     /number=16
     intron          12143. .12259
                     /number=16
     exon            12260. .12354
                     /number=17
     intron          12355. .12474
                     /number=17
     exon            12475. .12560
                     /number=18
     intron          12561. .12688
                     /number=18
     exon            12689. .12798
                     /number=19
     intron          12799. .13357
                     /number=19
     exon            13358. .14515
                     /number=20
BASE COUNT     3037 a    4168 c    4405 g    2905 t
ORIGIN


  Query Match             3.8%;  Score 37.4;  DB 3;  Length 14515;
  Best Local Similarity   58.6%;  Pred. No. 13;
  Matches   65;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy      54 CGGTGGCCCCGTTATTCGTCCAGGTGCCCAGGGAGGAGGACCCGCCTGCAGCATGAACCT 113
           ||  | || | | |  ||    |  |||| || ||  || |||  || |    |||  ||
```

```
Db    713 CGCCGCCCGCCTGAACCGCGACGCAGCCCTGGAAGACGGGCCCCGCTTCGAGGTGACGCT 772

Qy    114 GTGGCTCCTGGCCTGCCTGGTGGCCGGCTTCCTGGGAGCCTGGGCCCCCGC 164
          |  || |  | | |||| |  | | |||| |||||| |||  | || ||||
Db    773 GCTGCCCGAGCCATGCCGGACGCCGGGCTCCCTGGGGGCCGTGTCCTCCGC 823


RESULT  15
AF019858/c
LOCUS       AF019858      603 bp    DNA             PLN       04-AUG-1998
DEFINITION  Gynerium sagittatum internal transcribed spacer 1, 5.8S ribosomal
            RNA gene, and internal transcribed spacer 2, complete sequence.
ACCESSION   AF019858
NID         g3282481
VERSION     AF019858.1  GI:3282481
KEYWORDS    .
SOURCE      uva grass.
  ORGANISM  Gynerium sagittatum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Gynerium.
REFERENCE   1  (bases 1 to 603)
  AUTHORS   Hsiao,C., Jacobs,S.W.L., Barker,N.P. and Chatterton,N.J.
  TITLE     A molecular phylogeny of the subfamily Arundinoideae (Poaceae)
            based on the sequences of rDNA (ITS)
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 603)
  AUTHORS   Hsiao,C. and Chatterton,N.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-AUG-1997) Forage and Range Research Laboratory,
            USDA-Agriculture Research Service, Utah State University, Logan, UT
            84322-6300, USA
FEATURES             Location/Qualifiers
     source          1. .603
                     /organism="Gynerium sagittatum"
                     /db_xref="taxon:42053"
                     /tissue_type="leaves"
     misc_RNA        1. .223
                     /product="internal transcribed spacer 1"
     rRNA            224. .387
                     /product="5.8S ribosomal RNA"
     misc_RNA        388. .603
                     /product="internal transcribed spacer 2"
BASE COUNT       99 a    227 c    205 g     72 t
ORIGIN


  Query Match             3.8%;  Score 37.4;  DB 8;  Length 603;
  Best Local Similarity   55.9%;  Pred. No. 13;
  Matches   71;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy      3 CAGCTTGGCCTACAGCCCGGCGGGCATCAGCTCCCTTGACCCAGTGGATATCGGTGGCCC 62
          || |  | ||| | || ||| ||||   |   | | |  |||   ||  || |||||
Db    456 CACCGCGCCCTGCGGCGCGGGGGGCCAGACGCCGCGTCCCCCCCCGGGGATGGGTGGGAG 397

Qy     63 CGTTATTCGTCCAGGTGCCCAGGGAGGAGGACCCGCCTGCAGCATGAACCTGTGGCTCCT 122
          |||  || | |  |  ||||||| ||| |  ||| |   ||| | |   | |  ||  ||
Db    396 CGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCCCTCGGCCAGAAGGCCTCGGGCGCAACT 337

Qy    123 GGCCTGC 129
            || | |
Db    336 TGCGTTC 330


Search completed: July 24, 1999, 10:21:21
Job time: 11774 sec
```